Genomics and Genetic Engineering of *Helicoverpa armigera* Nucleopolyhedrovirus

Chen Xinwen

Promotoren: Prof. Dr. J.M. Vlak
Persoonlijk Hoogleraar bij de leerstoelgroep Virologie

Prof. Dr. R. W. Goldbach
Hoogleraar Virologie

Co-promotor: Prof. Dr. Hu ZhiHong
Director of the Wuhan Institute of Virology,
Chinese Academy of Sciences, Wuhan, P. R. China

Samenstelling promotiecommissie:

Prof. Dr. P. J. Krell (University of Guelph, Canada)
Prof. Dr. R. F. Hoekstra (Wageningen University, the Netherlands)
Dr. M. A. Huynen (European Molecular Biology Laboratory, Germany)
Dr. J. M. Sandbrink (Plant Research International, Wageningen UR, the Netherland)

GENOMICS AND GENETIC ENGINEERING OF *Helicoverpa armigera* NUCLEOPOLYHEDROVIRUS

Chen Xinwen

proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit
prof. dr.ir. L. Speelman,
in het openbaar te verdedigen
op woensdag 5 september 2001
des namiddags te 16:00 uur in de Aula

The work presented in this thesis was carried out at the Laboratory of Virology, Department of Plant Sciences, Research School 'Production Ecology and Resource Conservation', Wageningen University, Wageningen, the Netherlands, and at the Wuhan Institute of Virology, Chinese Academy of Science, financially supported by the Royal Academy of Science of the Netherlands (KNAW) and the Chinese Academy of Science (CAS).

Chen, X.

Genomics and genetic engineering of *Helicoverpa armigera* nucleopolyhedrovirus
Thesis Wageningen University- with ref. - with summary in English, Dutch and Chinese
ISBN 90-5808-421-3
Subject headings: cotton bollworm, baculovirus, biological control, molecular phylogeny

CONTENTS

Chapter 1

INTRODUCTION

COTTON CULTURE

Cotton is the most important natural textile fiber in the world. In China, the largest agricultural country in the world, cotton plays a significant role in economic and societal developments. It produces natural fiber, which forms the basis of the nation's textile industry, vegetable oil and animal food. Now about 4.0 million ha are grown with cotton annually, yielding a total production of 4.0 million tons. The three major cotton production regions are located along the Yangtze River, the Yellow River and in the Xinjiang Uygur Autonomous Region. At present, around 100 million farmers are involved in cotton production and 80 million people are employed by the cotton textile industry (Wang, 1996).

All cultivated cotton varieties in China belong to the species *Gossypium hirsutum* L. (upland cotton) with direct seeding, plastic mulching and seedling trans-planting at a density of about $7x10^4$ plants per ha. Monoculture and cotton-wheat inter-crop systems are common in China now. The latter system substantially increases productivity per unit land and has the added advantage of considerable reduction in insect damage at the seedling stage of cotton (Wang, 1990; Xia, 1997a).

Over 300 species of arthropod pests have been reported to be injurious to cotton, with 30 species of major significance. The major insect pest is the cotton bollworm (*Helicoverpa armigera* Hübner), which causes annual losses amounting to 20% of the attainable yield in China (CCRI, 1983). Severe damage due to cotton bollworm also reduced the area planted with cotton in China by 40% from the highest 6.7 million ha in the past to about 4.0 million ha in the present day (Xia, 1997b). The other key insect pests include cotton aphid (*Aphis gossypii* Glover), cutworm (*Agrotis ypsilon* Rott.), cotton thrips (*Thrips tabaci* Lindeman), red spider mite (*Tetranychus urticae* Koch), and pink bollworm (*Pectinophora gossypella* Saunders) (Xia, 1997a).

COTTON BOLLWORM (*Helicoverpa armigera*, HÜbner)

On a global scale, few insect pests cause as much economic crop loss as does *Helicoverpa armigera* and *H. zea* (Lepidoptera: Noctuidae). These species are widely distributed from the Pacific, Australia, through Southeast and southern Asia, the Middle East and southern Europe to Africa, with *H. armigera* predominantly distributed through the Old World, and

H. zea confined to the New World (Fitt, 1989). Typical of Noctuidae, *H. armigera* and *H. zea* are highly polyphagous, attacking a great variety of agricultural crops such as cotton, pepper, tomato, tobacco, maize, sorghum, sunflower, pigeonpea, chickpea, groundnut, soybean and okra (King, 1994). Damage is frequently located on the nitrogen-rich reproductive plant parts, such as young bolls, and thus influences yield directly.

As one of the key pests on crops in China, cotton bollworm (*H. armigera*) is widely distributed in cotton cultural areas from north to south, and from east to west (from 19^0N to 45^0N, from 124^0E to 75^0E), including the three major cotton-producing regions. In most regions the insect has four generations a year, while in the very south and west of China, such as Xiangjiang, there are 5 generations per year. Control of *H. armigera* on cotton has depended almost exclusively on chemical pesticides such as DDT, endosulfan and pyrethroids. However, *H. armigera* has a strong ability to develop resistance to chemical insecticides (McCaffery and Walker, 1991) which made chemical control unreliable and therefor there are increasing and urgent demands for bio-control or Integrated Pest Management (IPM) alternatives.

There has been much interest in classical biological control as a strategy in the management and control of *Helicoverpa* (King and Coleman, 1989, for review). One of the most frequently tried methods of achieving control with natural enemies has been by augmentative release of artificially reared parasites or predators (King and Coleman, 1989; Xia, 1997a). While the technical feasibility for controlling helothine bollworm populations in cotton by this method has been demonstrated, the results have not always

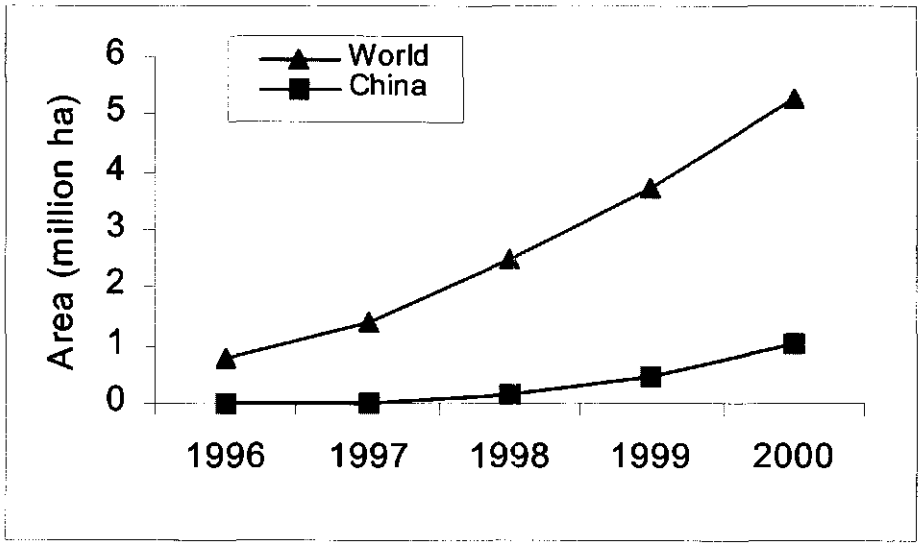


Figure 1-1. Global and Chinese area of transgenetic cotton

been consistent in the field (King and Coleman, 1989). Also, there has been long interest in exploiting entomophagous pathogens such as *Bacillus thuringiensis* (Bt) (Yearian *et al.*, 1980). Recently, Bt transgenic cotton variants against heliothines were developed and have been planted around the world (Figure 1-1) (James, 2000, for review). In China, the first commercialized Bt cotton was grown in 1997. The planting area was increased to 0.5 million ha (Figure 1-1) in 2000 (James, 2000). However, insect resistance on Bt cotton has been demonstrated and is of major concern (Gould *et al.*, 1997; Liu *et al.*, 1999; Shelton *et al.*, 2000). More information on the impact of transgenic cottons on the environment, sustainability, and global security is needed prior to wider use.

Baculoviruses have provided a solid alternative to chemical insecticides and have been developed to control bollworm and other cotton pests (Table 1-1) (Jones, 1994). For insect

Table 1-1. Reported field trails and use of baculoviruses for pest control on cotton (from Jones, 1994)

Insect	Virus	Country *
Agrotis segatum	*A. segatum GV*	Europe
Alabama agillacea	*A. agillacea NPV*	Brazil
	AcMNPV	USA
Amsacta spp.	*Amsacta* NPV	(Chad)
Anomis flava	*A. flava* NPV	(Australia), China, (Mali)
Bacculatrix thurberiella	AcMNPV	USA
Crythophlebia leucotreta	*C. leucotreta* GV	(Ivory Coast)
Diparopsis watersi	AcMNPV	Chad
Helicoverpa armigera	HaSNPV	Botswana, Chad, China, Egypt,India, Ivory Coast, S. Arfica, Thailand, Uganda, former USSR, Zimbabwe
	HzSNPV	Australia, Chad, Zimbabwe
	Mamestra brassicae NPV	Cameroon, Chad
Helicoverpa assulta	*M. brassicae* NPV	China
Helicoverpa punctigera	HzSNPV	Australia, USA
Helicoverpa virescens	HzSNPV	USA
	AcMNPV	USA
Helicoverpa zea	HzSNPV	USA, Nicaragua
Mythimna seperata	*M. seperata* NPV	China
Pectinophora gossypiella	AcMNPV	USA
Pseudoplusia includens	*P. includes* NPV	(Guatemala)
Spodoptera exigua	SeMNPV	(Chad), Thailand, USA
Spodoptera littoralis	*S. littoralis* NPV	Egypt, Israel
spodoptera litura	*S. littoralis* NPV	India, China
Spodoptera ornithogalli	*S. ornithogalli* NPV	USA
Trichoplusia ni	*T. ni* NPV	(Brazil), Columbia, USA
	AcMNPV	USA, Guatemala

* Countries in brackets = potential of control reported, field trails not undertaken.

species such as cotton bollworm that have developed resistance to chemical and/or *Bt* insecticides, the use of baculoviruses is one of the few options left for insect control (Brewer and Trumble, 1989). However, to meet the demands of the future, *i.e.* more effective control of the bollworm in cotton, baculoviruses with improved insecticidal properties are a potentially viable alternative. The work presented in this thesis focuses on the baculovirus, *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus (HaSNPV). After a detailed molecular analysis of HaSNPV, recombinant HaSNPV with improved insecticidal properties will be made, which could serve as alternative agents for cotton bollworm control in China.

BACULOVIRUSES

Taxonomy

Baculoviruses form a diverse group of parasites pathogenic only for arthropods (Federici, 1999). While the vast majority of baculoviruses has been isolated from over 600 insect species, primarily from Lepidoptera, but also from Hymenoptera and Diptera and several have been found in crustacean order Decapoda. Based on the number of virions occluded in the occlusion body, baculoviruses are divided into two taxonomic genera: *Nucleopolyhedrovirus* (NPV) and *Granulovirus* (GV). Many NPV virions are occluded within a proteinic matrix called the occlusion body or polyhedra, while, one virion of GV or, rarely, two or more, is occluded in a matrix called granule. Nucleocapsid envelopment of occluded virions occurs either within the nucleus (NPV) or in the nuclear-cytoplasmic milieu after rupture of the nuclear membrane (GV). NPVs have two morphotypes, SNPV and MNPV, depending on the single (S) or multiple (M) packaging of the nucleocapsids in the virions (Figure 1-2). The aggregation of nucleocapsids within the envelopes does not appear to be phylogenetically significant, but is only of taxonomic importance (Hu, 1998; Blissard *et al.*; 2000, this thesis). On the other hand, phylogenetic analysis based on several viral genes indicated that NPVs can be divided into two groups: group I and II (Zanotto *et al.*, 1993; Bulach *et al.*, 1999). Group I contains the relatively well characterized baculoviruses, like *Autographa californica* (Ac) MNPV, *Orgyia pseudotsugata* (OpMNPV) and *Bombyx mori* (Bm) NPV, while *Lymantria dispar* (Ld) MNPV, *Spodoptera exigua* (Se) MNPV and *Helicoverpa zea* (Hz) SNPV are typically claded into group II (Hu *et al.*, 1998). *H. armigera* SNPV (HaSNPV or HearNPV) is also a member of group II.

Infection cycle

Transmission of a baculoviruses in an insect population occurs via ingestion of food, via soil contaminated with occlusion bodies by insect larvae or via ovipostion of eggs contaminated with occlusion bodies (Blissard *et al.*, 2000). Infection is initiated when the occlusion bodies are dissolved in the alkaline juices in the insect midgut. The dissolution is aided by the presence of alkaline proteinases in the midgut lumen of most Lepidoptera.

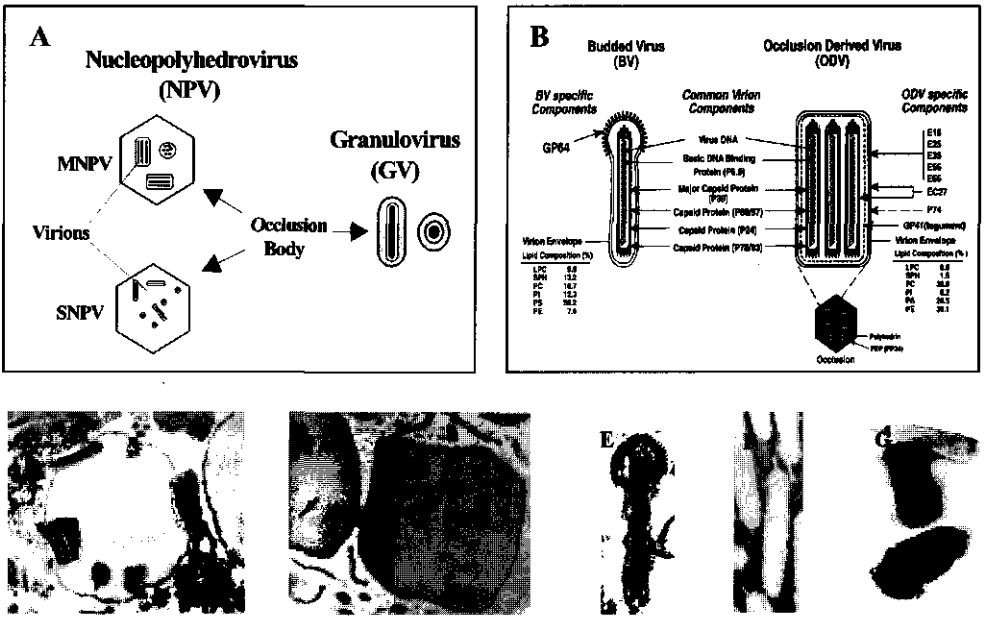


Figure 1-2. Baculovirus occlusion bodies, virions, and nucleocapsid forms. (Top) A: The structures of occlusion bodies from baculoviruses in the genera *nucleopolyhedrovirus* (NPV) and *granulovirus* (GV) are illustrated. Virions embedded in nucleopolyhedrovirus occlusion bodies may contain multiple nucleocapsids (MNPV) or single nucleocapsids (SNPV). B: The two baculovirus virion phenotypes are illustrated as diagrams with shared and phenotype-specific components. (Bottom) Transmission electron micrographs of occlusion bodies of the MNPV (*Helicoverpa armigera* MNPV, C) and SNPV (*H. armigera* SNPV, D) types. Transmission electron micrographs of the BV (*Lymantria dispar* MNPV, E) and ODV (HaSNPV, F and HaMNPV, G). Notably, HaSNPV ODV contains a small tail at one end (F). Adopted from Blissard *et al* (Blissard *et al.*, 2000). Electron micrographs (C, D, F, G) are a courtesy of G. Zhang (Wuhan, China)

The virions released from the occlusion bodies - occlusion-derived virus (ODV) - pass through the peritrophic membrane and fuse with the midgut epithelial cell plasma membrane. The nucleocapsids are released into the cytoplasm and migrate to the nucleus where transcription of viral genes and replication of the viral genome takes place. The nucleocapsids synthesized in the nucleus pass through the nuclear membrane, bud from the cell and acquire a new envelope from the plasmalemma to become budded virions (BV). ODVs and BVs have similar nucleocapsids but are morphological distinct and have specific polypeptides (Funk *et al.*, 1997, for review, and Fig. 1-1). The BVs produced in epithelial cells of the midgut spread via the hemolymph (Granados and Lawler, 1981) and the tracheal system (Engelhard *et al.*, 1994) into all tissues of the insect causing a secondary infection, but the predominant target is the fat body cell. Occlusion of virions does not occur in midgut epithelial cells as they are sloughed off in to the gut

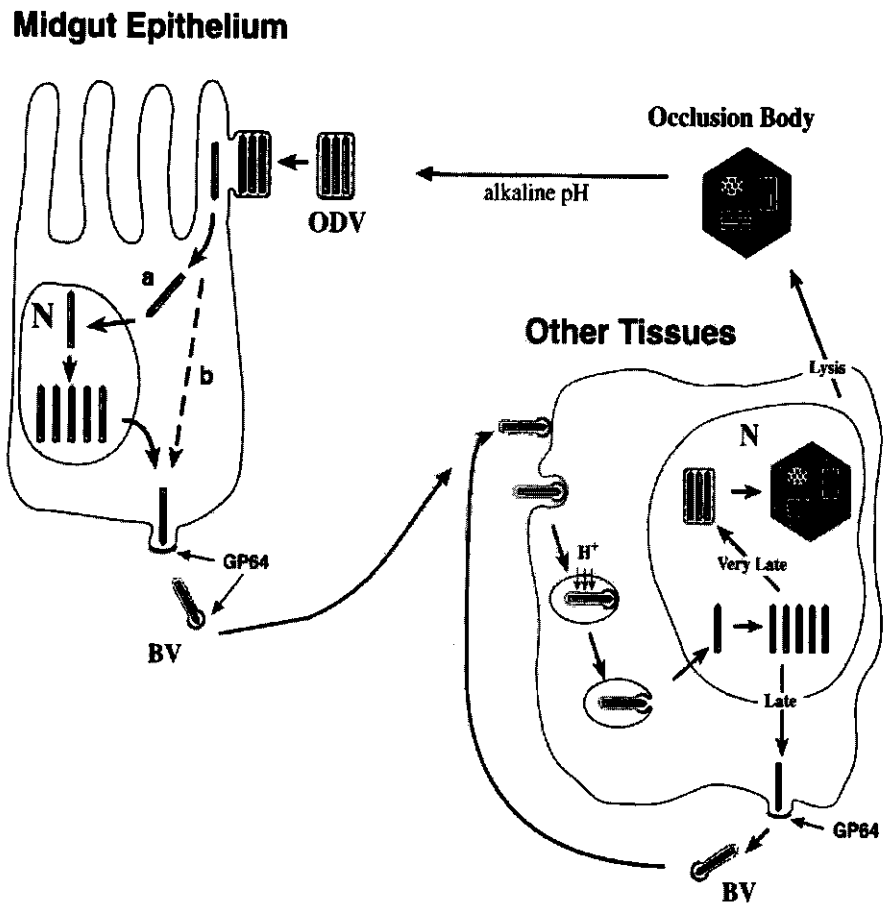


Figure 1-3. The baculovirus infection cycle (Blissard, 1996). On the left, the infection of midgut epithelial cells by the occlusion derived virus (ODV) is illustrated. After fusion of the ODV at the plasma membrane, two possible routes of nucleocapsid migration are indicated (a and b). In what is thought to be the typical route (a), nucleocapsids (NCs) are transported to the nucleus where gene expression, DNA replication, and assembly of progeny NCs occur. Newly assembled NCs then migrate from the nucleus towards the plasma membrane. A portion of incoming NCs may bypass the nucleus (b), and migrate directly to the basolateral plasma membrane. NCs bud from the basal side of the epithelial cell at sites where the major envelope fusion glycoprotein (EFP) GP64 of the BV phenotype (indicated by arrows), has accumulated. The right side of the figure represents the infection of insect cells, other than midgut epithelial, by budded virus (BV) phenotype. After BV binding and uptake into the endosome, the acidification of the endosome triggers fusion of the viral and endosome membrane, after which the NCs are released into the cytoplasm. Although the modes of entry of the ODV and BV are different and specific for each virus phenotype, once NCs are released into the cytoplasm, progression of infection appears similar. Except for midgut epithelial cells at the very late stage of the infection cycle nucleocapsids become embedded within proteinaceous occlusion bodies (OBs) in the nucleus. Upon death and liquefaction of the host, OBs are released into the environment to initiate a new round of infection.

lumen and replaced by regenerating cells in most lepidopterous insects. Virion occlusion does occur in rare occasions in midgut epithelial cells (Flipsen *et al.*, 1993; Flipsen *et al.*, 1995b).

In the early stage of the secondary infection, cells produce BVs that spread from cell to cell within the insect tissues and virion occlusion takes place at the late stages of infection. At the end of the infection cycle, the cells and tissues disintegrate and occlusion bodies are released into the environment. Virus encoded proteins such as fibrillin, chitinase and cathepsin have been reported to be involved in this process (van Oers and Vlak, 1997; Hawtin *et al.*, 1995; Hawtin *et al.*, 1997; Slack *et al.*, 1995; Hill *et al.*, 1995). The next infection cycle starts when occlusion bodies are ingested by other susceptible insects (Figure 1-2).

Genome organization

Baculoviruses contain a circular, supercoiled double-stranded DNA genome with a size ranging from 80-180 kilobasepairs (kb) depending on the species. Recently, DNA sequence analysis has begun to reveal the distinctive feature of the baculovirus genomes and the extent of their diversity. So far, five NPV and two GV genome sequences have been reported. These are AcMNPV (Ayres *et al.*, 1994), OpMNPV (Ahrens *et al.*, 1997), LdMNPV (Kuzio *et al.* 1999), BmNPV (Gomi *et al.*, 1999), SeMNPV (IJkel *et al.*, 1999) and *Xestia c-nigrum* (Xc) GV (Hayakawa *et al.*, 1999) and *Plutella xylostella* (Px) GV (Hashimoto *et al.*, 2000). The sizes of these genomes range from 101 to 179 kb with G+C contents of 40-48%, and 120 to 181 potential open reading frames (ORF) in the different genomes were found (Table 9-1). The ORFs, in general, are tightly packed with minimal intergenic regions and their orientations are almost evenly distributed along the genome. Both strands of the genome are involved in coding functions and the genes are generally unspliced. Gene classes (early, late and very late genes, see further discussion) are not clustered in baculoviral genomes. One unique feature is the presence of several homologous regions with repeated sequences (*hrs*), which might be involved in the regulation of DNA replication and gene expression (transcription). Gene duplications are also frequently observed in baculoviral genomes.

Genomic comparisons by GeneParityPlot analysis indicated that the genomes of group I baculoviruses such as AcMNPV, BmNPV and OpMNPV basically have a similar gene content. Some inversions and insertions (deletions) have, however, been observed. SeMNPV and LdMNPV, which belong to group II, might have a recent common ancestor and are more distantly related to group I (Hu *et al.*, 1998; IJkel *et al.*, 1999). The GV group is quite separate from the NPV group. However the gene organization in the interest "central" part of the baculoviral genomes is highly conserved and is a key feature in the alignment of baculovirus genomes (IJkel *et al.*, 1999, Heldens *et al.*, 1998). However, Group II NPVs are largely unresolved due to the lack of sequenced genomes, hence the in

sequencing the genome of HaSNPV.

Gene expression

Baculovirus gene expression is organized in a sequential, cascade-like fashion in which each successive phase is dependent on the previous one (Blissard and Rohrmann, 1990). Regulation of baculoviral gene expression occurs at the transcriptional level. Transcription of baculovirus genes is temporally regulated, and three main classes of genes are recognized: early, late and very late. Early genes are transcribed by host RNA polymerase II and contain typical eukaryotic consensus transcription motifs such as a TATA box and a CAGT motif. The products of early genes, among which are powerful transregulators such as immediate early gene 1 (*ie*-1), function to prepare the host cell for virus replication and to accelerate replicate events (Friesen, 1997). Late genes are transcribed by a viral encoded α-amanitin-resistant RNA polymerase starting at a consensus A/GTAAG motif. Many late genes encoding viral structural proteins, such as *vp*39 and *p*6.9, the major capsid proteins and the basic core protein, are abundantly transcribed, whereas very late genes encoding the occlusion-related proteins like polyhedrin or granulin are transcribed after the late genes (Vlak and Rohrmann, 1985). Another abundantly expressed very late gene is *p*10, which encodes a small polypeptide that effects nuclear disintegration in the final phases of cell death (van Oers and Vlak, 1997, for review). These two very late genes (*ph* and *p*10) are expressed from very strong viral promoters that are extensively utilized in baculovirus insect cell expression systems (Smith *et al.*, 1983; Martens *et al.*, 1995).

Transient early and late gene transcription and DNA replication assays suggested that four virus-encoded proteins, immediate early transregulators 0, 1, 2 (IE0, IE1, IE2) and PE38, regulate early gene transcription (Friesen, 1997, for review), while approximately 19 virus-encoded proteins known as late expression factors (LEFs) are necessary for late gene transcription (Lu and Miller, 1997, for review). DNA replication is required for late gene transcription. Of the approximately 19 LEFs, 9-10 appear to be involved in DNA replication and 5 of which *helicase*, *ie*-1, *lef*-1, *lef*-2 and *lef*-3, are essential for DNA replication (Lu *et al.*, 1997, for review).

The field of baculovirology has been developing rapidly in the last two decades. One reason of this advance was that baculoviruses form an attractive alternative to chemical pesticides for insect pest control (Moscardi, 1999; Cunningham, 1998). Due to their unique features of replication and gene expression and the availability of suitable cell lines, baculoviruses are also widely used as an eukaryotic expression vectors for the production of recombinant proteins of biotechnological or pharmaceutical importance (Possee, 1997, Kost & Condreay 1999, for review). The technological advancement associated with the generation of expression vectors has also been successfully used to develop enhanced baculovirus pesticide agents (Bonning and Hammock, 1996). Recently, baculovirus–derived vectors have emerged as a possible tool for gene transfer into

mammalian cells and holds a promise in gene therapy (Hofmann *et al.*, 1995; Boyce and Bucher, 1995; Condreay *et al.*, 1999; Sarkis *et al.*, 2000). An in-depth analysis of the baculovirus genome organization, replication and gene expression strategy is a prerequisite for optimal exploitation of these economically important applications. The genetic analysis also leads to a better understanding of baculovirus replication and pathology.

BACULOVIRUSES IN PEST CONTROL

The abundant use of hazardous chemical pesticides has caused increased resistance of many insect pests to the commonly used chemical pesticides and has also caused considerable environmental pollution. Alternative bio-control strategies, such as the use of parasites, predators and microorganisms, are strongly being promoted. Baculoviruses are a

Table 1-2. Baculoviruses registered as microbial insecticides

Insect Pest	Baculovirus	Crops	Country and regions
Adoxophyes orana	GV	apple	Switzerland
Agrotis segetum	GV	carrots, cotton	Denmark, Russia
Anagrapha falcifera	NPV	cotton, vegetable	USA
Anticarsia gemmatalis	NPV	soybean	Brazil
Autographa californica	NPV	cotton, vegetable	USA, China
Cydia pomonella	GV	apple, pears	Europe, USA, Russia
Dendrolimus sibericus	NPV	pine forest	Russia
Helicoverpa spp	NPV	cotton, vegetables	USA, China, Russia, Thailand, Australia
Hyphantria cunea	NPV & GV	orchards & parks	Bulgaria, Russia
Leucoma salicis	NPV	parks	Russia, Poland
Lymantria dispar	NPV	forests	USA, Canada, Russia
Mamestra brassicae	NPV	cabbage, pea, sugar beet	Europe, Bulgaria, Russia
Malacosoma neustria	NPV	orchards & parks	Russia
Neodiprion lecontei	NPV	forests	Canada
Neodiprion sertifer	NPV	pine	Europe, Poland, Russia
Orgyia pseudotsugata	NPV	forests	USA, Canada
Oryctes rhinoceros	NPV	coconut palm	Indian
Plodia interpunctella	GV	stored nuts	USA
Pieris rapae	GV	vegetables	China
Spodoptera albula	NPV	cotton, cabbage, melon	Guatemala
Spodoptera litura	NPV	vegetables	China
Spodoptera litoralis	NPV	cotton	Africa, France
Spodoptera exigua	NPV	vegetables, ornamentals	USA, Europe, Thailand
Spodoptera frugiperda	NPV	corn	Brazil
Trchoplusia ni	NPV	cotton and vegetables	USA

natural component of the ecosystem and have been successfully applied as biological control agents in IPM.

Some baculoviruses have a host range restricted to one or a few related insect species and can cause epizootics in nature that appear to regulate the size of the host insect population. The infectivity, specificity and safety to non-target organisms make them attractive biological agents to replace chemical insecticides (FAO/WHO, 1973). Nowadays, baculoviruses are employed worldwide as ecologically sound biological pesticides to control a variety of pest insects (Table 1-2) (Cunningham, 1995, for review). The earliest commercial venture with a viral insecticide goes back to the 1960s with the development of Virion H (HzSNPV), which later (1976) was replaced to Elkar, for the control of *Heliothis spp* in North America (Greer *et al.*, 1971; Ignoffo, 1973; Ignoffo and Couch, 1981; Shieh, 1989). This product was accompanied by three non-commercial preparations produced by the US Forestry Service, namely Gypcheck (LdMNPV), TMBioControl-1 (OpMNPV) and Neocheck-S [*Neodiprion sertifer* (Nese) NPV]. Successful examples of baculoviruses as insecticides are SeMNPV against beet army worm on cotton, flowers and vegetables (Smits and Vlak, 1994; Kolodny-Hirsh and Dimmock, 1996), LdMNPV against Douglas fir tussock moth in forests, *Anticarsia gemmatalis* (Ag) MNPV against soybean caterpillar in Brazil (Moscardi, 1999), and HaSNPV against cotton bollworm in China as well as in other countries such as Thailand, Australia, India, South Africa, Egypt, Zimbabwe, and the former USSR (Zhang, 1994; Ketunuti and Prathomrut, 1989; McKinley, 1971; 2000; for review: Jones, 1994; Table 1–3).

Since the 1970s baculoviruses have been tested and used to control insect pests in China. So far, more than twenty baculoviruses have been developed as biological agents for insect pest control (Hu, 1998). Five of which, HaSNPV, AcMNPV, *Spodoptera litura* (Splt) MNPV, *Gynaephora ruoergensis* (Gyru) NPV and *Plutella xylostella* (Px) GV, have been registered as biological pesticides and are mass-produced. The most successful case is HaSNPV that is used on approximately 100,000 ha annually to control the heliothinae, including *H. armigera, H. assulta,* and *H. zea,* on cotton, tobacco, and hot pepper (Zhang, 1989; Zhang, 1994).

GENETIC ENGINEERING OF BACULOVIRUS FOR PEST CONTROL

The application of baculoviruses as microbiological agents has not met the optimal potential to control the pest in crops, forests, and pastures. One of the major limitations is their relative slow speed of action. Typically, a baculovirus infection takes about one week to kill the insect hosts. During this time the insects can still cause serious damage to the crop. Secondly, baculoviruses are far less virulent in later instar insects, which cause the majority of damage in crops (Evans, 1981). Thirdly, baculoviruses often have a relatively narrow host range. Finally, baculoviruses are not very persistent in the environment as

they are rapidly inactivated by sunlight (UV). Through genetic engineering, the required insecticidal properties, such as increased speed of action, enhanced virulence and extended host range, might be achieved (Vlak, 1993b).

Different strategies have been tried to improve the baculovirus properties, mainly focused on increasing the speed of action. These approaches, technologically, are either deletion of a baculovirus gene, which normally function to extend the lifetime of the infected insect,

Table 1-3. Summary of recombinant baculoviruses developed for pest control

Virus designation*	Heterologous protein	Host	Speed improvement	Reference
gene-deletion				
VEGTDEL	*egt*- deletion	Sf	22%	O'Reilly & Miller, 1991
VEGTDEL	*egt*- deletion	Tn	0.5-1 day	Treacy *et al.*, 1997
VEGTDEL(LdMNPV)	*egt*- deletion	Ld	33%	Slavicek *et al.*, 1999
Gene insertion (insect hormone)				
BmDH5(BmNPV)	MS diuretic hormone	Bm	~1days	Maeda,1989
AcRP23(B)JHE	Hv JHE	Tn	No#	Hammock *et al.*, 1990
AcUW(2) JHE	Hv JHE	Tn	No	Bonning *et al.*, 1992
Ac-JHE29	Hv JHE with R for K substitution	Tn Hv	No	Bonning *et al.*, 1995
Ac-JHE524	Hv JHE with R for K substitution	Tn Hv	No	Bonning *et al.*, 1995
VWTPTTHM, vEGTPTTHM	Bm PTTH, *egt*-deletion	Sf	No	O'Reilly *et al.*, 1995
ie1-JHE-KK *p10*-JHE-KK	Hv JHE with R for K substitution	Hv	No	Jarvis *et al.*, 1996
Ac-JHE-KK	Hv JHE with R for K substitution	Tn	4-9%	Kunimi *et al.*, 1996
Ac-JHE-KK	Hv JHE with R for K substitution	*Pi*	No	Kunimi *et al.*, 1997
AcJHE-KSK AcJHE-KHK	Hv JHE with animo acids substitution	Hv	No#	Van Meer *et al.*, 2000
Gene insertion (toxin gene)				
BmAaIT(BmNPV)	Toxin AaIT from *A. Australis*	Bm	~40%	Maeda *et al.*, 1991
AcST3	Toxin AaIT from *A. Australis*	Tn	24%	Stewart *et al.*, 1991
AcUW2(B)AaIT	Toxin AaIT from *A. Australis*	HV	36%	McCutchen *et al.*, 1991
AcMNPVAaIT	Toxin AaIT from *A. australis*	Tn	20-30%	Kunimi *et al.*, 1996
p10-AaIT	Toxin AaIT from *A. australis* alternate promoters	Hv	*P10*:22% *Ie1*:10%	Jarvis *et al.*, 1996
vSP-tox-34	Neurotoxin 34 from *P. tritici*	Tn	39%	Tomalski & Miller, 1991
vSP-tox21A	Neurotoxin 34 from *P. tritici*	Tn	49%	Tomalski *et al.*, 1993a
vSP-toxin34, vp6.9toxin34, vDA26toxin34, vHSPtoxin34	Neurotoxin 34 from P. *tritici*, various single sequence (ss),various promoter (pr)	Tn Sf	Ss:26-47%, pr:28-58%; ss:47-53%, pr:39-59%	Lu *et al.*, 1996
vSP-TOX34#4	Neurotoxin 34#4 from *P. tritici*	Hv Tn	27-33% 56%	Watkins *et al.*, 1997
vEV-HA5fl7	Venom Dol m V gene *Delichovespula maculata*	Tn	No	Tomalski *et al.*, 1993b
vSAt2p+	Neurotoxin As II from *Anemonia sulcata*	Tn Sf	Tn:37% Sf:36%	Prikhod'ko *et al.*, 1996
vMAg4+	Neurotoxin Aga-IV from *Angelenopsis aperta*	Tn Sf	Tn: 17% Sf: 43%	Prikhod'ko *et al.*, 1996
vSSh1p+	Neurotoxin Sh I from *Stichodactyla helianthus*	Tn Sf	Tn:37% Sf:40%	Prikhod'ko *et al.*, 1996

vAcTaITX1	Toxin from spider *Tegenaria agrestis*	Tn Se	Tn:20% Se:18%	Hughes *et al.*, 1997
AcNPVLIT	alph-latroinsectotoxin from black window spider	Hv Tn	1:4%; 2:12%	Watkins *et al.*, 1997
AcLqhIT2	Toxin LqhIT2 from *Leiurus quinquestriatus hebreus*	Tn Se Hv	50/56% 37/47%, 40/45%	Presnail *et al.*, unpublished
AcLIT1.p10 AcLIT2.pol	Toxin LqhIT1 and LqhIT2 from *Leiurus quinquestriatus hebreus*	Ha	32%; 24%	Gershburg *et al.*, 1998
Ac(PH+)BT	delat-endotoxin	Tn	No	Merryweather *et al.*, 1990
AcBtm, AcBt5, AcBt3, AcBt5/3	Cryl(b) & truncated forms	Hv	No	Ribeiro & Crook, 1993
AcMNPV/JM2, AcMNPV/FW3	Truncated CrylA(b)	Se	No	Martens *et al.*, 1995
Gene insertion (other genes)				
BV13T, BV13.3940	Maize mitochondrial geneURF13	Tn	~40%	Korth & Levings, 1993
AcMNPV.chi	Mc chintase gene	Sf	22-23%	Gopalakrishnan *et al.*, 1995
vhcf-1z, vhcf-1del	AcMNPV host cell factor 1	Tn Sf	+ 20-29%, no	Lu & Miller, 1996
AcMNPV-MycAs	Human c-myc antisense	Sf	28%	Lee *et al.*, 1997
AcBX-PBAN-4	Hz PBAN	Tn	19-26%	Ma *et al.*, 1998
AcMNPV-	JHE antisense	Hv	No	Hajos *et al.*, 1999
Deletion and insertion				
vJHEEGTD	HV JHE, *egt*-deletion	Tn	No	Eldridge *et al.*, 1992a
VEHEGTD	Ms eclosion hormone, *egt*-deletion	Sf	No	Eldridge *et al.*, 1992b
vSPTox34	Neurotoxin 34 from *Pyemotes tritici*	Tn	~2 days	Tomalski & Miller, 1992
HzEGTp6.9tox34, HzEGThsptox34, HzEGTDA26tox34 (HzSNPV)	Neurotoxin 34 from mite *P. tritic*, *egt*-deletion, various promoter (pr)	Hz	35-42% 34-47%	Popham *et al.*, 1997

Note. ET: effective time, LT: lethal time; ST: survial time; LC: lethal concentration
Bm: *Bombyx mori*; Hv: *Heliothis virescens*; Sf: *Spodoptera exigua*; Tn: *Trichoplusia ni*;
Ha: *Helicoverpa armigera*
* Freeding reduction
The parental virus other than AcMNPV indicated in the bracket

or incorporation of genes, whose expression in the target insect interferes with some critical aspect of insect physiology and hopefully results in feeding arrest or death of the insect (Table 1-3, Bonning & Hammock, 1996; Black *et al.*, 1997, for review). However, over- expression of insect genes that normally regulate key aspects of insect physiology or development, such as diuretic (Maeda, 1989), prothoracicotropic (O'Reilly *et al.*, 1995) and eclosion hormone (Eldridge *et al.*, 1992b) and juvenile hormone esterase (JHE) (Hammock *et al.*, 1990) did not significantly increase the baculovirus speed of action nor reduce the food consumption of the infected insects. Along with the deletion of *egt* gene from the genome to enhance baculoviruses efficacy, insertion of insect-specific neurotoxins holds the greatest promise for delivering a commercially viable baculovirus insecticide (O'Reilly and Miller, 1991; Stewart *et al.*, 1991; Tomalski and Miller, 1991; Prikhod'ko *et al.*, 1996; Popham *et al.*, 1997; Gershburg *et al.*, 1998; Table 1-3).

Baculovirus encode an EGT that specifically inactive the ecdysteroid hormone by conjugating it with glucose or galactose. Previous results with different baculoviruses indicated that deleting *egt* from the genomes could increase the viral killing speed up to 30% and cause a significant reduction of the amount of food consumed (O'Reilly and

Miller, 1991; Slavicek *et al.*, 1999). Expression of insect-selective toxins in the baculovirus system has proved to be highly successful for increasing virus efficacy in insect-pest control. The insect-selective toxin derived from the scorpion *Buthus eupeus* (*Be*IT) was first introduced into AcMNPV (Carbonell *et al.*, 1988). However, for unknown reasons, this toxin did not increase the efficacy of the recombinant virus. Also expression of the toxins from *Bacillus thuringiensis* did not improve baculovirus pesticidal properties (Martens *et al.*, 1990; Merryweather *et al.*, 1990). Expression of the insect-specific toxin *Aa*IT from the North Africa scorpion *Androctonus australis* in baculovirus produced the most encouraging results (Stewart *et al.*, 1991, McCutchen *et al.*, 1991). This toxin acts on the neuronal sodium channel, causing presynaptic excitatory effects. Larvae infected with these recombinant baculoviruses exhibited dorsal arching and increased irritability and decreased feeding. Lethal times are reduced by 25-40% compared with those of the wild-type virus, and the feeding damage by larvae infected with the recombinant virus was reduced by 50% compared with damage caused by larvae infected with wild-type viruses. The other toxins which also enhanced recombinant baculovirus insecticidal efficacy include TxP1 from the straw itch mite, *Pyemotes tritici* (Tomarski & Miller, 1991; 1992), LqhIT1 and LqhIT 2, scorpion depressant toxin from the scorpion *Leiurus quinquestriatus hebreus* (Gershburg *et al.*, 1998).

The effectiveness of the genetically modified baculoviruses has been tested in the field and have shown their potential to enhance crop protection (Cory *et al.*, 1994; Gard, 1997; Hernandez-Crespo *et al.*, 1999) The field trial results carried out by American Cyanamid indicated that AcMNPVs with an *egt* deletion would kill target insect species about 15-20% faster than wild type. An AcMNPV with an *egt*-deletion and AaIT-expression had the fastest killing speed, which resulted in an increased level of control over the wild type and the *egt*-deletion forms (Black *et al.*, 1997). Thus recombinant baculovirus hold possible commercial potential.

Helicoverpa armigera NUCLEOPOLYHEDROVIRUS (HaSNPV)

Two types of *H. armigera* NPVs, single nucleocapsid nucleopolyhedrovirus (HaSNPV or HearNPV) and multi-nucleocapsid nucleopolyhedrovirus (HaMNPV) were isolated in China (JSM & CCNU, 1975; Jiang *et al.*, 1978; WIV-CAS, 1979). Due to its high virulence, HaSNPV has been adopted for mass production as a viral pesticide and has been widely used to control the insect pests in China and in other countries (Jones, 1994). The morphology and pathology of HaSNPV has been described to some degree (WIV-CAS, 1979; Xie *et al.*, 1980; Zhang *et al.*, 1983a; 1983c; Lua and Reid, 2000). Apart from its phenotype, HaSNPV appears to have a host range restricted to members of the genus *Helicoverpa* including *H. zea, H. virescens* and *H. assulta* (Ignoffo and Couch, 1981; Sun and Zhang, 1994). By comparison, AcMNPV infects more than 70 insect species and SeMNPV can only replicate in its original host. Morphogenesis of HaSNPV, *e.g.*

formation of the polyhedral envelope, is different from that of AcMNPV (Xie *et al.*, 1980; Lua and Reid, 2000).

Besides HaSNPV, baculoviruses have also been reported from other heliothine insects, including *H. zea* (HzSNPV) (Ignoffo, 1965a, b, c), *H. virescens* (HeviNPV) (Ignoffo, 1965a), and *H. punctigera* (HepuNPV) (Teakle, 1973). Those viruses are either SNPVs or MNPVs. Biochemical, serological and molecular studies revealed that SNPVs and MNPVs of *Heliothis/Helicoverpa* are quite different, while the various SNPVs have a high degree of similarity (McCarthy *et al.*, 1978; Getting and McCarthy, 1982). *Helicoverpa* SNPVs containing different genotypes share, more or less, the same host range but show differences in virulence and restriction enzyme profiles (Hughes, 1982; Hamm, 1982; Sun and Zhang, 1994).

THE MAIN AIMS OF THE PROJECTS AND OUTLINE OF THE THESIS

Insight into the molecular characteristics that specify the biological properties of HaSNPV is important not only to understand the basis of host range and virulence, but also for the successful and bio-safe genetic modification of this virus. Despite the extensive usage of HaSNPV/HzSNPV in pest control over many decades, the molecular knowledge of the HaSNPV genome is limited to viral DNA restriction enzymes profile analysis (Getting and McCarthy, 1982; Knell and Summers, 1984; Corsaro and Fraser, 1987; Sun and Zhang, 1994). Therefore, understanding key steps in the infection process such as virus entry, gene expression, genome replication and virion assembly is important prior to the development of genetic engineering strategies for HaSNPV. Furthermore, no SNPV has been sequenced to date and it would be interesting to see if this morphology has a taxonomic basis. The major aim of the present study was to investigate the genome organization, gene structure and function of HaSNPV and finally to generate recombinant HaSNPV with enhanced insecticidal properties using a targeted approach.

HaSNPV and HzSNPV are chosen in this project not only because they have been widely used in pest control and their potential commercial value as recombinant insecticides, but also for some of their unique features such as the single nucleocapsid (SNPV) phenotype and a narrow host range. In comparison to MNPVs, knowledge of the molecular biology and genetics of SNPVs is very limited. The study of the HaSNPV/HzSNPV genomes will allow comparison of SNPVs to MNPVs at the gene and the genome levels and might also reveal the molecular basis of baculovirus morphotype. It has been reported that HaSNPV and HzSNPV share a high degree of sequence homology, have similar biological features including the morphology and pathology. A detailed analysis of their genomes could reveal their relationship and evolution.

One of the unique features of baculoviruses is the presence of polyhedra, which are large

round to cubical shaped proteinoaceous structures that contain the ODVs. Polyhedra, most likely, have evolved to protect the virions outside the host against their decay due to environmental factors or upon disintegration of tissues. They also deliver the virus to midgut cells that are susceptible to infection. Polyhedrin, the major component of occlusion bodies, has been the subject of detailed investigations because it is structurally and functionally highly conserved (Vlak and Rohrmann, 1985, Rohrmann, 1986). The expression of the polyhedrin gene is under the control of a powerful late promoter, which has been widely used to express foreign genes in baculovirus-insect expression systems or to generate recombinant viral pesticides. Identification of the polyhedrin gene, consequently, is the first step in the analysis of the HaSNPV genome (**Chapter 2**). This analysis produced the first molecular evidence that HaSNPV and HzSNPV might be different isolates of the same virus species.

Baculoviruses encode an ecdysteroid UDP-glucosyltransferase (EGT) that abrogates the insect host regulation of metamorphosis by conjugating ecdysteroids with sugars and hence blocking molting/pupation of the insect. Deletion of *egt* enhances baculovirus killing speed. Insects infected with an *egt*-null virus also consume less foliage than those infected with the wild type (O'Reilly and Miller, 1991). To achieve *egt*-deletion with HaSNPV to generate a recombinant with improved pesticidal properties, the identification and sequence analysis and phylogeny of the HaSNPV *egt* is described in **Chapter 3**.

Egt, an auxiliary gene, is not essential for virus replication *in vitro*, but, nonetheless, provides baculovirus with some selective advantage in nature. Compared to other essential genes, auxiliary genes might have a different kind of selection pressure in the evolutionary history and may have different evolutionary patterns. To further analyze the phylogenetic relationship between HaSNPV and other baculoviruses, an essential gene, late expression factor 2 (*lef-2*), of HaSNPV was characterized and compared to other baculoviruses (**Chapter 4**). *Lef-2* is putatively assigned as a primase-processivity factor and involved in viral DNA replication and transcription regulation (Kool *et al.*, 1994; 1995). Also, the topological structure of the *lef-2* tree is in accordance with that for *egt*. Comparison of the relative *lef-2* locations in the genomes also indicates that gene order might be an independent parameter in baculovirus phylogenetic analysis.

Construction of a plasmid library and establishment of a detailed map were essential for molecular studies and genetic engineering. A physical map has been constructed by using restriction enzyme analysis and hybridization. This map was further refined by sequencing the plasmid library either in both directions or only at the ends of the inserts (**Chapter 5**). These sequence data facilitated an initial view of the gene content and organization of the HaSNPV genome. Both gene content and GeneParityPlot analysis confirmed the phylogenetic position of HaSNPV in baculovirus Group II, but also indicated that the HaSNPV genome has a distinct unique gene arrangement. High identity and overall

homology between HaSNPV and HzSNPV suggested that these two viruses might be two variants of the same virus species.

So far, the nucleotide sequences of five MNPVs (Ayres *et al.*, 1994; Gomi *et al.*, 1999; Ahrens *et al.*, 1997; Kuzio *et al.*, 1999; IJkel *et al.*, 1999) and two GVs (Hayakawa *et al.*, 1999; Hashimoto *et al.*, 2000) have been reported. However, no SNPV genome has been sequenced to date and it is therefore of prime interest to see whether the sequence of HaSNPV would reveal unique features contributing, among others, to the SNPV phenotype and to the specificity of this virus for heliothine insects. The entire HaSNPV genome was sequenced and delineated in **Chapter 6**.

HaSNPV and HzSNPV share a high degree of sequence identity in their ORFs (**Chapter 1, 6, 7**). They also have the same heliothine host range (Hughes *et al.*, 1983; Hamm, 1982; Sun and Zhang, 1994). In order to reveal if they are different isolates of the same virus, the entire genome sequence of HzSNPV was determined (**Chapter 7**). The sequence data indicated that the two viruses have similar gene content and arrangement. Alignment of the genome sequences indicated that the conserved genes have a lower nucleotide substitution rate, the homologous regions (*hrs*) are the highest variable regions, and that the two *bro* genes in both viruses might have a different evolutionary history.

Since the aim is to enhance the efficacy of HaSNPV as a pesticide, both deletion of the *egt* gene and insertion of the AaIT gene were tested **(Chapter 8)**. These recombinant HaSNPVs expressing AaIT gene and no longer produce EGT proved to have better insecticidal properties. Furthermore green fluorescent protein (GFP) marker gene was introduced to select the recombinant *in vitro* and to follow its fate in the environment.

Chapter 2

NUCLEOTIDE SEQUENCE ANALYSIS OF THE POLYHEDRIN GENE OF *Helicoverpa armigera* SINGLE NUCLEOCAPSID NUCLEOPOLYHEDROVIRUS

SUMMARY

The polyhedrin gene of the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus (HaSNPV) was localized by Southern blot hybridization with a probe from the *Autographa californica* MNPV (AcMNPV) polyhedrin gene. Nucleotide sequencing showed that the open reading frame (ORF) of the HaSNPV polyhedrin gene was 738 nucleotides (nt) long and encoded a protein of 246 amino acids with a predicted size of 29 kD. Alignment with the other polyhedrin sequences in the GenBank indicated the highest homology with the polyhedrin gene of *H. zea* SNPV (HzSNPV). There are four different nucleotides and only one change (nt 197 C → A) resulting in an amino acid change in HaSNPV (Thr→Lys). This change did not alter the secondary structure of the protein molecule. The promoter regions of both polyhedrin genes were identical including the location of the ATAAG core motif, where the transcription starts. The amino acid homology of HaSNPV polyhedrin with other SNPV and MNPV polyhedrins was much lower. The results indicate that HaSNPV and HzSNPV are probably variants of a single baculovirus species.

This chapter has been published largely as

Chen, X., Hu, Z. and J.M. Vlak. 1997, Nucleotide sequence analysis of the polyhedrin gene of *Heliothis armigera* single nucleocapsid nucleopolyhedrosis virus. *Virologica Sinica* 12: 346-353.

Chapter 2

INTRODUCTION

The bollworm *Helicoverpa armigera* is a major pest insect in cotton, which causes annually about 20% yield loss of the cotton culture in China. A single nucleocapsid nucleopolyhedrovirus was isolated from *H. armigera* (HaSNPV or HearNPV) in the early 1970s and has been extensively used to control this insect in cotton as well as in vegetables, such as hot pepper. In 1993 HaSNPV was registered as the first viral pesticide in China and is commercially produced ever since. The morphology and pathology of HaSNPV have been studied in some detail (Xie *et al.*, 1980; Zhang *et al.*, 1983a), but molecular biological studies were limited only to the analysis of viral genomic DNA with restriction enzymes. These studies indicated that the genome size is about 130 kb (Sun and Zhang, 1994).

At the end of the infection, baculoviruses produce proteinaceous occlusion bodies (OB), which contain the rod-shaped virions with a single (S) nucleocapsid or multiple (M) nucleocapsid per virion (Blissard *et al.*, 2000). The major component of baculovirus OBs is the polyhedrin protein. This protein has a size of about 30 kD and seems highly conserved among baculoviruses (Vlak and Rohrmann, 1985). OBs protect virions in the environment against decay. Synthesis of polyhedrin in infected cells is temporally regulated and under the control of a powerful very late promoter. This property has led to the development of the successful baculovirus-insect cell expression system (King and Possee, 1992; O'Reilly *et al.*, 1992) and the generation of baculovirus insecticides with improved insecticidal properties (Vlak, 1993a).

Quite a few polyhedrin genes have been reported from different baculoviruses, including *Autographa californica* (Ac) MNPV (Hooft van Iddekinge *et al.*, 1983); *Bombyx mori* (Bm) NPV (Gomi *et al.*, 1999), *Orgyia pseudotsugata* (Op) MNPV (Leisy *et al.*, 1986a), OpSNPV (Leisy *et al.*, 1986b), *H. zea* (Hz) SNPV (Cowan *et al.*, 1994), *Buzura suppressaria* (Busu) NPV (Hu *et al.*, 1993), *Spodoptera exigua* (Se) MNPV (van Strien *et al.*, 1992) and *Cydia pomonella* (Cp) GV (Crook *et al.*, 1997) (Rohrmann, 1986; Hu, 1998, for review). Most polyhedrin genes were reported from MNPVs and only a few from SNPVs. Here we reported the sequence and analysis of the polyhedrin gene from HaSNPV and compared this sequence with other baculovirus polyhedrins.

MATERIALS AND METHODS

Virus and DNA

HaSNPV was propagated in fourth instar *H. armigera* larvae by contamination of the artificial diet with polyhedra. The occlusion bodies (OB) were purified by differential and rate zonal centrifugation (Sun and Zhang, 1994). The DNA was isolated directly from purified OB after alkaline treatment, followed by proteinase K and SDS treatment,

phenol/chloroform extraction and dialysis (King and Possee, 1992). The purity and the concentration of the DNA were determined spectrophotometrically.

Southern blot hybridization

Two primers, VW90 (GGAGATCTAATATGCCGGATTACTCA) and VW91 (GGAGATC-TTTAATACGCCGGACCAGT) containing the flanking sequences of the AcMNPV polyhedrin gene were designed. The AcMNPV polyhedrin gene was synthesized and labeled with ^{32}P-dATP by PCR with VW90 and VW91 as primers and *Eco*RI-I fragment of AcMNPV as the template (Vlak and Smith, 1982). Viral DNA was digested with restriction enzymes according to the manufacturer protocols (Life-Techologies/BRL), separated in 0.7% agarose and transferred to nitrocellulose (Hybond) filters. The blots were hybridized with ^{32}P-labeled AcMNPV polyhedrin gene probe according to Southern (1975) under non-stringent conditions.

Molecular cloning, sequencing and computer analysis

The bands, which hybridized with the AcMNPV polyhedrin gene sequences, were recovered from the agarose gel and were ligated into pTZ19R, digested with appropriate restriction enzymes. The local region was mapped by analysis of the positive clones with restriction enzymes. The selected positive clones were sequenced with universal M13 and specific primers by using the dideoynucleotide chain-terminating method of Sanger *et al.* (1977).

Sequences were analyzed with the aid of the UWGCG (Version 10) computer program; DNA and deduced amino acids sequences were compared with data in GenBank.

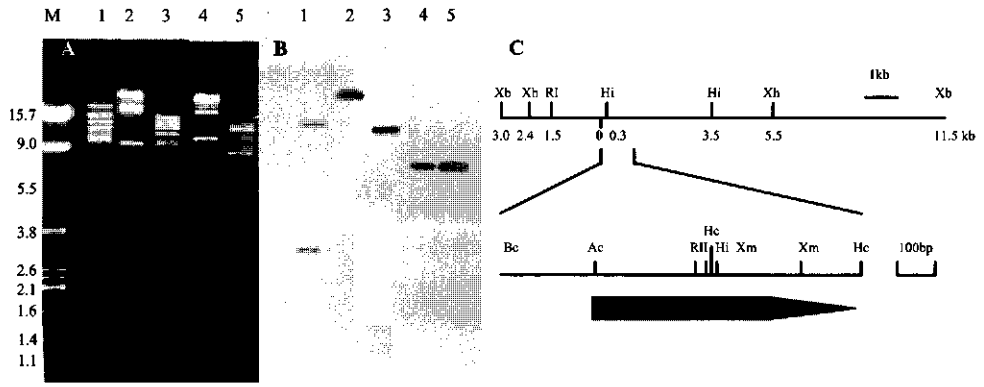


Figure 2-1. HaSNPV DNA digested with *Hin*dIII, *Sst*I, *Xba*I and *Xho*I (Lane 1, 2, 3, 4) and AcMNPV DNA digested with *Eco*RI (lane 5); and Southern hybridization (B) under non-stringent conditions with a ^{32}P-labeled PCR product of the AcMNPV polyhedrin gene as a probe. Physical map of HaSNPV *Xba*I-A fragment that contains the polyhedrin gene (C). The arrow shows the HaSNPV polyhedrin gene and its direction of transcription. Ac = *Acc*I, Bc = *Bcc*I, RI = *Eco*RI, Hi = *Hin*dIII, Hc = *Hin*cII, Xb = *Xba*I, Xm = *Xmn*I, Xh = *Xho*I.

RESULTS AND DISCUSSION

The HaSNPV polyhedrin gene was located on a *Sst*I - A (26.5 kb), a *Xba*I - A (14.5 kb) and a *Xho*I - F (7.2 kb) fragment, respectively, while two *Hin*dIII fragments, D (13.3 kb) and J (3.2 kb), hybridized with AcMNPV polyhedrin gene (Fig. 2-1A, 2-1B). All these fragments were then cloned into pTZ19R. A detailed physical map of this region was also constructed (Fig. 2-1C). The *Hin*dIII-D and J fragments were first sequenced from the both termini and the polyhedrin was thus determined (Fig. 2-2). The double stranded sequence of the polyhedrin gene was generated from the 7.2 kb *Xho*I-F fragment, which contains the entire polyhedrin gene, by using synthesized primers.

The polyhedrin gene is 738 nucleotides long with the potential to encode a 246 amino acids-long protein with a predicted size of 29 KD. This size agrees well with the estimate of the polyhedrin protein from HaSNPV OBs by SDS-PAGE (30 KD, Sun *et al.* 1998). Nucleotide sequence alignment indicated that the polyhedrin gene of HaSNPV has an extremely high degree of identity with that of HzSNPV (Table 1 and Fig. 2-2). Four nucleotides, located at nucleotide residues 197 (C/A), 549 (C/T), 612 (T/C) and 672 (A/G), were found different in HaSNPV and HzSNPV, respectively. Only one nucleotide change (nt 197 C → A) resulted in an amino acid change in HaSNPV (Thr → Lys) (Fig. 2-2) (Cowan *et al.*, 1994).

The promoter region of the HaSNPV polyhedrin gene was found to be identical to that of the HaSNPV polyhedrin gene and similar to those of other baculovirus (Fig. 2-3). In the 5' untranslated region (UTR), a putative baculovirus late transcription motif is present in HaSNPV with the canonical core sequence TAAG (Vlak and Rohrmann, 1985) starting at nucleotide position 51 upstream of the translation start condon ATG. One conserved sequence, TTGTGA, presented in all baculovirus 5' UTRs was detected. (O'Reilly *et al.*, 1992). In general, the promoter region of the HaSNPV polyhedrin gene had higher identity to similar regions of SNPVs (BusuNPV and OpSNPV) than to those of MNPVs (Hu *et al.*, 1993; Leisy *et al.*, 1996). One typical poly(A) signal sequence, AATAAA, was found at nucleotide position 157 downstream of the translation stop codon. The AAT_9 motif found in SeMNPV and some of other MNPVs (Van Strien *et al.*, 1992) is absent in HaSNPV. Future experiments will characterize the transcription of the HaSNPV polyhedrin gene.

The amino acid sequence of HaSNPV polyhedrin was compared to six other known NPV polyhedrins (Table 1, Fig. 2-4). Besides HzSNPV polyhedrin, in which only one amino acid was found different to the HaSNPV polyhedrin, the polyhedrins of the other two SNPVs, BusuNPV and OpSNPV, are also more closely related to HaSNPV polyhedrin than to the MNPV polyhedrins. The average amino acid identity among baculovirus polyhedrins is 84%, indicating that the polyhedrin gene is one of the most conserved genes

```
          1 Met Tyr Thr Arg Tyr Ser Tyr Ser Pro Thr Leu Gly Lys Thr Tyr Val
 HaSNPV   1 ATG TAT ACT CGT TAC AGT TAC AGC CCT ACT TTG GGC AAA ACC TAT GTG
HzSNPV      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

         17 Tyr Asp Asn Lys Tyr Phe Lys Asn Leu Gly Ala Val Ile Lys Asn Ala
HaSNPV   49 TAC GAC AAC AAA TAC TTT AAG AAT TTA GGT GCT GTT ATT AAA AAT GCC
HzSNPV      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

         33 Asn Arg Lys Lys His Leu Glu Glu His Glu His Glu Glu Arg Asn Leu
HaSNPV   97 AAA CGC AAG AAG CAT TTA GAG GAG CAC GAA CAT GAA GAA CGC AAC TTG
HzSNPV      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

         49 Asp Ser Leu Asp Lys Tyr Leu Val Ala Glu Asp Pro Phe Leu Gly Pro
HaSNPV 145 GAT TCG CTC GAC AAA TAC TTG GTG GCG GAA GAT CCT TTT TTG GGA CCT
HzSNPV      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

         65 Gly Thr Asn Gln Lys Leu Thr Leu Phe Lys Glu Ile Arg Ser Val Lys
HaSNPV 193 GGC ACA AAT CAA AAA CTA ACT TTG TTT AAA GAG ATT CGC AGC GTT AAG
HzSNPV      ..  .A. ... ... ... ... ... ... ... ... ... ... ... ... ... ...
               Lys

         81 Pro Asp Thr Met Lys Leu Val Val Asn Arg Ser Gly Arg Glu Phe Leu
HaSNPV 241 CCC GAC ACA ATG AAG CTT GTA GTT AAC TGG AGC GGT CGC GAA TTT CTT
HzSNPV      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
                       HindIII

         97 Arg Glu Thr Arg Thr Arg Phe Met Glu Asp Ser Phe Pro Ile Val Asn
HaSNPV 289 CGC GAA ACT TGG ACG CGT TTC ATG GAA GAC AGT TTT CCC ATT GTA AAC
HzSNPV      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

        113 Asp Gln Glu Ile Met Asp Val Phe Leu Ser Val Asn Met Arg Pro Thr
HaSNPV 337 GAC CAA GAA ATT ATG GAC GTG TTT CTG TCT GTT AAT ATG CGA CCA ACC
HzSNPV      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

        129 Lys Pro Asn Arg Cys Tyr Arg Phe Leu Ala Gln His Ala Leu Arg Cys
HaSNPV 385 AAA CCG AAC CGT TGT TAC CGA TTC TTA GCG CAA CAC GCT CTG CGT TGT
HzSNPV      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

        145 Asp Pro Asp Tyr Ile Pro His Glu Val Ile Arg Ile Val Glu Pro Ser
HaSNPV 433 GAT CCC GAC TAT ATT CCT CAC GAA GTC ATT CGT ATT GTA GAA CCT TCC
HzSNPV      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

        161 Tyr Val Gly Ser Asn Asn Glu Tyr Arg Ile Ser Leu Ala Lys Lys Tyr
HaSNPV 481 TAT GTA GGC AGT AAC AAC GAG TAC AGA ATT AGT TTA GCC AAA AAA TAC
HzSNPV      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

        177 Gly Gly Cys Pro Val Met Asn Leu His Ala Glu Tyr Thr Asn Ser Phe
HaSNPV 529 GGC GGT TGT CCC GTT ATG AAC TTG CAC GCT GAA TAC ACT AAT TCC TTT
HzSNPV      ... ... ... ... ... ... ..T ... ... ... ... ... ... ... ... ...

        193 Glu Asp Phe Ile Thr Asn Val Ile Arg Glu Asn Phe Tyr Lys Pro Ile
HaSNPV 577 GAA GAT TTC ATT ACC AAC GTA ATT TGG GAG AAC TTT TAC AAA CCA ATT
HzSNPV      ... ... ... ... ... ... ... ... ... ... ... ..C ... ... ... ...

        209 Val Tyr Val Gly Thr Asp Ser Ala Glu Glu Glu Glu Ile Leu Leu Glu
HaSNPV 625 GTT TAC GTA GGC ACT GAT TCT GCC GAA GAA GAG GAA ATA CTC CTA GAA
HzSNPV      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G

        225 Val Ser Leu Ile Phe Lys Ile Lys Glu Phe Ala Pro Asp Ala Pro Leu
HaSNPV 673 GTT TCT TTG ATA TTT AAG ATC AAA GAA TTT GCA CCT GAC GCT CCG CTA
HzSNPV      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

        241 Tyr Thr Gly Pro Ala Tyr ***
HaSNPV 721 TAC ACT GGT CCT GCA TAT TAA
HaSNPV      ... ... ... ... ... ... ...
```

Figure 2-2. Alignment of HaSNPV and HzSNPV polyhedrin genes. The predicted amino acids are indicated. Dots represent identical nucleotides. The *Hin*dIII restriction site separating *Hin*dIII D and J is indicated.

```
HaSNPV  TAAA ATAAG TATTTTTTTCC TATTGTTCAA GATTGTGAAA AATCAAAT ATCCC ATA
HzSNPV  TAAA ATAAG TATTTTTTTCC TATTGTTCAA GATTGTGAAA AATCAAAT ATCCC ATA
BusuNPV TTCA ATAAG TATTTTTTTCC TATTGTAAAA CATTGCGAAA AATCAAAT ACAAC ATA
OpSNPV  CTCA ATAAG TATTTTTGTCC TTTCGTAAAA CATTGTGAAA TTTCAAAT ACACC ATA

MbMNPV  AAAT GTAAG TAATTTTCTCC TTTCGTAGAA GATTGTGAAA AATAAAAT       ATA
SfMNPV  AATT GTAAG TAATTTTTTCC TTTCGTAAAA CATTGTGAAA AAATAAAT       ATA
SeMNPV  AATT GTAAG TAATTTTTTCC TTTCGTAAAA CATTGTGAAA AAATAAAT       ATA
OpMNPV   TTA ATAAG TAATTTCCTGT TATTGTAACA ATTTTGTAAA AAAATTTC CTATAACC
AcMNPV  TTAA ATAAG TATTTTACTGT TTTCGTAACA GTTTTGTAAT AAAAAAAC CTATAA ATA
```

Figure 2-3. Comparison of the promoter region of nine NPV polyhedrin genes. The core motif, which contains the start of transcription, is in bold. The ATG translation start is the next to the right of the sequence.

Table 2-1. Amino acid and nucleotide sequence identities (%) of seven NPV polyhedrins

	HzSNPV		BusuNPV		OpSNPV		MbMNPV		SeMNPV		AcMNPV	
HaSNPV	**99**	99	**90**	79	**89**	78	**86**	75	**86**	75	**85**	73
HzSNPV			**91**	79	**89**	79	**87**	76	**87**	75	**86**	73
BusuNPV					**95**	81	**94**	81	**91**	79	**90**	77
OpSNPV							**94**	79	**91**	79	**88**	78
MbMNPV									**92**	82	**88**	80
SeMNPV											**85**	76

* Bold denotes amino acid sequence identities.

of baculoviruses. A nuclear localization sequence KRKK was found at amino acid 33-36 of the HaSNPV polyhedrin (Jarvis *et al.*, 1992).

Morphologically, NPVs are subdivided into two subgroups: SNPV and MNPV based the number of nucleocapsids in the virions. Here we found that the HaSNPV polyhedrin gene and its 5' UTR showed a higher degree of identity to those of three other SNPVs, suggesting that these SNPVs have a more close relatedness. However, phylogenetic analysis of all baculovirus polyhedrins failed to separate SNPVs from MNPVs to form an independent group (Hu, 1998).

HaSNPV and HzSNPV shared a very high identity on nucleotide (99%) and amino acid (99%) sequences. This suggests that HaSNPV and HzSNPV might be two variants of the same virus species. This agrees with the observation of their biological features. These two viruses have the same host range. Both can infect *H. armigera, H. zea, H. virescence* and *H. assulta* (Sun *et al.*, 1998). Moreover, HaSNPV and HzSNPV genomic DNAs have a similar restriction enzyme profile. Southern analysis also shows a high overall identity (Sun *et al.*, 1998). Detailed molecular studies will reveal, whether HaSNPV and HzSNPV

```
HaSNPV  :  .......S.............................E...H...N..S.............  :  62
HzSNPV  :  .......S.............................E...H...N..S.............  :  62
BusuNPV :  .......K.............................E.............................  :  62
OpSNPV  :  ....................................Q....A..H.....................  :  62
MbMNPV  :  ........................S.....N....Y.............................  :  62
SeMNPV  :  .........A..................S........E................E..........  :  62
AcMNPV  :  .-PD...R.............................FA......A......N.............  :  61

                *
HaSNPV  :  ...T...........S.....................................S..  :  124
HzSNPV  :  ...K...........S.....................................S..  :  124
BusuNPV :  ....................................................................  :  124
OpSNPV  :  ....................................................................  :  124
MbMNPV  :  ....................................................................  :  124
SeMNPV  :  ..............I.....................................................  :  124
AcMNPV  :  ........................G.K....Y.....................................  :  123

HaSNPV  :  ..........................................................Y.........A  :  186
HzSNPV  :  ..........................................................Y.........A  :  186
BusuNPV :  ....................................................................  :  186
OpSNPV  :  ......................E.............................................A  :  186
MbMNPV  :  ....................................................................  :  186
SeMNPV  :  ..................................V.....................................  :  186
AcMNPV  :  ........................D.....................................  :  185

HaSNPV  :  .......D..TN.....................................................  :  246
HzSNPV  :  .......D..TN.....................................................  :  246
BusuNPV :  ....................................................................  :  246
OpSNPV  :  ............H.....................................................  :  246
MbMNPV  :  ..........................................................N.....  :  246
SeMNPV  :  ..........................G.....................N.....  :  246
AcMNPV  :  ....................................................................  :  245
```

Figure 2-4. Amino acid alignment of seven baculovirus polyhedrins. The one-letter code designation for amino acids is used. The dots denote identical amino acids. The asterisk indicates the only different amino acid between HaSNPV and HzSNPV.

are variants of the same virus species and also may give additional information on the organization of SNPV genomes in comparison with those of MNPVs and GVs.

Chapter 3

ANALYSIS OF THE ECDYSTEROID UDP-GLUCOSYLTRANSFERASE GENE OF *Helicoverpa armigera* SINGLE-NUCLEOCAPSID BACULOVIRUS

SUMMARY

An ecdysteroid UDP-glucosyltransferase (*egt*) gene was identified from the single (S) nucleocapsid nucleopolyhedrovirus of *Helicoverpa armigera* (HaSNPV). In baculovirus-infected insects the viral enzyme (EGT) plays a pivotal role in abrogating the insect molting process. The open reading frame of the *egt* gene is 1545 nucleotides long, encoding a putative protein of 515 amino acids with a M_r of 59.1. The 5'-noncoding region contains a putative early (CAGT) and late (TAAG) motif for transcription initiation, a transcription enhancer sequence (CGTCGC) and two TATA boxes. A putative polyA signal, AATAAA, was found downstream of the translation stop codon. A putative signal peptide of 21 residues was present at the N-terminus of the EGT. The HaSNPV *egt* gene has a high degree of nucleotide and amino acid sequence homology to the *egt* genes of *Buzura suppressaria* SNPV and *Spodoptera exigua* MNPV. The HaSNPV EGT shares ten conserved motifs with other EGTs. A phylogenetic tree of twelve baculovirus EGTs was constructed by using maximum parsimony analysis, suggesting that SNPVs do not form a separate clade within the baculovirus family.

This chapter has been published as

Chen, X., Hu, Z., Jehle, J. A., Zhang, Y., and J. M. Vlak 1997. Analysis of the ecdysteroid UDP-glucosyltransferase gene of *Heliothis armigera* single-nucleocapsid baculovirus. *Virus Genes* 15: 219-225.

INTRODUCTION

Baculoviruses are large DNA viruses mainly occurring in insects. These viruses are categorized into two taxonomic groups: the *nucleopolyhedroviruses* (NPV) and the *granuloviruses* (GV) (Murphy *et al.*, 1995). In NPVs the rod-shaped nucleocapsids are packaged singly (S) or in multiples (M) in virions that are in turn occluded into large proteinaceous capsules, called occlusion bodies or polyhedra. Baculoviruses are widely used as biological control agents of insect pests in agriculture and forestry.

Baculoviruses contain a gene (*egt*) coding for the enzyme ecdysteroid UDP-glucosyltransferase (EGT) (see O'Reilly, 1995, for review). In infected insect larvae this viral enzyme conjugates glucose or galactose to ecdysteroids and this results in a delay or in the absence of insect molting (O'Reilly and Miller, 1989). As a consequence, infected insects continue to grow and feed, and produce high amounts of occlusion bodies (polyhedra). Deletion of the *egt* gene from baculoviruses results in normal and timely molts of infected insects and in reduced polyhedra yield since these insects die at a smaller size. Insects infected with *egt* deletion mutants also die faster, possibly through an indirect effect on the functioning of the Malpighian tubules (O'Reilly and Miller, 1991; Flipsen *et al.*, 1995a).

A single-nucleocapsid nucleopolyhedrovirus (SNPV) has been isolated from *H. armigera* (Lepidoptera, Noctuidae) in the People's Republic of China. This virus is named HearNPV/HaSNPV (Murphy *et al.*, 1995), but is also known as HaSNPV (Sun and Zhang, 1994). Preparations of HaSNPV have been successfully used in the last decade to control *H. armigera* in Chinese cotton (Zhang, 1989; Zhang, 1994). However, improvements of its insecticidal properties, in particular its speed of action, are sought through genetic engineering (Bonning and Hammock, 1996; Hu and Vlak, 1997). In this paper we describe the location and sequence analysis of the *egt* gene of HaSNPV as a prelude to the engineering of *egt* deletion mutants of this virus.

The double stranded DNA genome of HaSNPV has been characterized by restriction enzyme analysis and appeared to be similar to *Helicoverpa zea* SNPV (Sun and Zhang, 1994). Recently, the polyhedrin gene of HaSNPV has been identified and sequenced (Chen *et al.*, 1997b) and a physical map of the viral DNA is being prepared (Chen *et al.*, in preparation). The nucleotide sequence of the HaSNPV polyhedrin gene was very similar to that of HzSNPV suggesting that HzSNPV and HaSNPV are variants of a single genotype (Cowan *et al.*, 1994; Chen *et al.*, 1997b).

METHODS

Viruses

The HaSNPV was isolated from diseased *H. armigera* larvae found in the Hubei province of the People's Republic of China and produced in fourth instar larvae (Zhang, 1989). The virus was purified by sucrose gradient centrifugation and stored at -20° C until use. Viral DNA was isolated from polyhedra according to Sun and Zhang (1994) and digested with restriction enzymes according to the manufacturer protocols (Life Technologies/BRL).

Southern hybridizations

*Hin*dIII and *Sst*I digests were separated in 0.7% agarose and electrotransferred to nitrocellulose (Hybond) filters. The blots were hybridized with a ^{32}P-labelled *Pst*I-*Kpn*I fragment of the BusuNPV *egt* gene (Hu *et al.*, 1997) according to Southern (1975).

Cloning and sequencing

A HaSNPV *Hin*dIII library was constructed by shotgun cloning of *Hin*dIII-digested HaSNPV DNA into pTZ19. The termini of selected *Hin*dIII fragments were sequenced by automatic sequencing using universal pTZ primers. Further sequencing was done with *Hin*dIII or *Sst*I (sub) clones using universal and specific oligonucleotide primers. The sequences were screened against databases (GenBank) and searched for sequence homology. Cloning procedures were as described in Sambrook *et al.* (1989).

Phylogenetic analysis

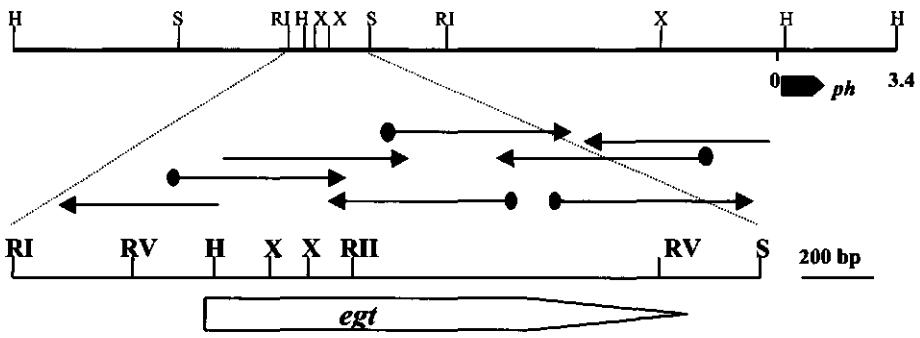
The alignment of EGTs was carried out by PileUp of the UWGCG computer program (Devereux *et al.*, 1984, release 8.0). A most parsimonious tree of EGTs was constructed by using the branch-and-bound algorithm of PAUP 3.1 (Swofford, 1993). Bootstrap analysis (100 replicates) was performed by using a heuristic search algorithm.

RESULTS AND DISCUSSION

The HaSNPV *egt* gene was identified by hybridization of an *egt* probe of *Buzura suppressaria* SNPV (Hu *et al.*, 1997) to a *Hin*dIII and *Sst*I restriction digest of HaSNPV DNA. This probe consisted of a *Pst*I-*Kpn*I fragment of 747 bp (nt 576-1323) of the coding sequence of BusuNPV *egt* containing the conserved EGT domains II and III (O'Reilly, 1995; Fig. 3-2) and hybridized to a 12.8 kb *Hin*dIII fragment (*Hin*dIII-D) and a 4.0 kb *Sst*I fragment (*Sst*I-H) (data not shown). This *Sst*I-H fragment, in turn, hybridized to *Hin*dIII-D as well as to *Hin*dIII-I indicating that the HaSNPV *egt* gene overlaps with these two fragments (Fig. 3-1a). *Hin*dIII-D also contained part of the polyhedrin gene at one end of this fragment (Chen *et al.*, 1997b). A partial physical map was constructed linking *Hin*dIII-K to *Hin*dIII-D and I (Fig. 3-1a). Sequencing of the termini using universal and specific primers and comparison with databases identified 114 nucleotides (nt) of the 5'

end of the *egt* gene on *Hin*dIII-I and 1431 nt of the 3' end on *Hin*dIII-D (Fig. 3-1b).

A



B

*Eco*RV
TAGATATCTGTAATACACATCAAAGTAAACATGTTTACTTAAACAGTAGCTGAATAATAATTTATAACATAGCGACGCCACTATAR
GATGCAGCATCCGTCCGTTGGTCATCTTTCGATAGACGCTCTGACCCATAAACGGACGTGCGCTAATTTTTTTTATTGCTAAATTCAAA
ATGTACAAACAGATAATAACTATGTTATTGTTGGTGTTGTTTCTGTCGGTTCTGGATGGAGCGCGTATCCTGTGCGTTTTTCCTGTTCCT
M Y K Q I I T M L I L V L F L S V L D G A R I L C V F P V P
*Hin*dIII
TCGTACAGTCATCATGCAGTGTTCGAAGCTTACACCAATGCTCTAGCGTCGCGTGGCCATACAATAGTCAGAATTACACCGTTTCCCACT
S Y S H H A V F E A Y T N A L A S R G H T I V R I T P F P T
AAGAAAAACGATTCATCCAACGTGACAGATGTCGACGTTAGCTTGTCGAAAGATTATTTTAAAAGTCTTGTGGACCGATCTAGACTGTTC
K K N D S S N V T D V D V S L S K D Y F K S L V D R S R L F
AAGAAACGAGGCGTTATTTCGGAAACGTCCAGCGTGACCGCTCGCAATTACATCAGTCTAGTACACATGTTGATTGATCAATTCTCTATG
K K R G V I S E T S S V T A R N Y I S L V H M L I D Q F S M
GAGAGTGTACGACAATTGATCGAATCTAACAATGTTTTCGATTTGTTGGTGACCGAAGCCTTTCTAGATTATCCTCTGGTGTTTTCGCAT
E S V R Q L I E S N N V F D L L V T E A F L D Y P L V F S H
TTGTTTGGCGATGTGCCTGTCATACAAATTTCGTCGGGTCACGCTTTGGCCGAAAATTTTGAGACAATGGGAGCCGTGAGCCGACATCCC
L F G D V P V I Q I S S G H A L A E N F E T M G A V S R H P
ATTTACTATCCAAATTTGTGGCGCAACAAATTTCAAAATTTAAACGTTTGGGAGATAATAACGGAAATCTATACAGAACTGGTGCTGTAC
I Y Y P N L W R N K F Q N L N V W E I I T E I Y T E L V L Y
TTGGAATTTGCTCTGTTTAGCCGACGAACAAACTAAAATGCTTCGCCATCAATTCGGACCAAACACGCCCAGCGTGGAAGAACTGCGACAA
L E F A R L A D E Q T K M L R H Q F G P N T P S V E E L R Q
CGCGTTCAATTATTGTTTGTGAATACGCATCCGCTGTTTGATAATAACAGACCAGTACCGCCGAGTGTACAATATTTGGGAAGTCTACAT
R V Q L L F V N T H P L F D N N R P V P P S V Q Y L G S L H
CTTGATCGAAACAATGATATTAACGAACAGCAAACGATGGACTATAATTTGATGCAATTTTTAAATAATTCTACAAACGGTGTGGTGTAC
L D R N N D I N E Q Q T M D Y N L M Q F L N N S T N G V V Y
GTGAGCTTCGGTACGTCTATACGAGTTTCAGACATGGACGACGAATTTCTGTTTGAATTTATAACAGCTTTCAAGCAATTACCCTATAAT
V S F G T S I R V S D M D D E F L F E F I T A F K Q L P Y N
ATATTGTGGAAGACCGATGGAATGCCCATGGAACACGTGCTGCCTAAAAATGTGTTGACACAAACTTGGCTGCCGCAACACCATGTATTG
I L W K T D G M P M E H V L P K N V L T Q T W L P Q H H V L
AAACACAGCAATGTAGTTGCTTTTGTTACTCCAGGCGGAATGCAGTCAACGGACGAAGCCATCGACGCTTGTGTACCACTAATCGGAATC
K H S N V V A F V T P G G M Q S T D E A I D A C V P L I G I
CCGTTTATGGGCGACCAAGCATACAATACCAATAAATACGAAGAACTCGGAATCGGACGCAACCTCGATCCCGTAACGCTCACAAGTCAT
P F M G D Q A Y N T M K Y E E L G I G R N L D P V T L T S H
ATTTTGGTGTCTGCCGTTTTAGATGTGACCGTCAACAACAAGAGTCGCTACACAGATAATATTAAAGCATTGAATCGTTCCACTAATTAT
I L V S A V L D V T V N N K S R Y T D N I K A L N R S T N Y
CGAACACGGAAACCTATGGAAAAGCCATCTGGTACACAGAACATGTAATTGATAATGGTAAAAATCCCATTTTAAAAACGAAGCCGCC
R T R K P M E K A I W Y T E H V I D N G K N P I L K T K A A
*Eco*RV
AACGTATCGTATAGCAAATATTATATGAGTGATATCATCGTTCCTGCTATAACGTTTTGGTAATGACTCATTTGGGTCAGGCTATTCGG
N V S Y S K Y Y M S D I I V P A I T F L V M T H L G Q A I R
CGGTTGGTTGTTATTTAATACTGTATGACAATGTACACATGTGTTAATAAAAAAGGCATTACTAATATTTAGATTGTTTCAAATTATTTA
R L V V I *
CGCATGACTACCCGTCTCCTATTGCGCAGCTACGCTAGCTTTAAATACAGCCGATGGCGTAGTAAAGTTCATTTAAATATCTAAATTGGT
TAGTTCAACATCGCGGTGCGAGCGCACGACTTATACCATGCATCGGTCCAATAGTAACAGCAGCAAATACAAACAATCGCTGATAAATCG
CTTTGAACTGGAATACAAAAGTGTGTCTGTGCGCGATTTGCAAAAATTGTCAGCGGCCATGTATCGTTTGTTGGCTGTGAACGATAAACT
TATGGAAAATTTACAAACTCTACCGATGCATTATAGAGCTCGA

Figure 3-1. A, Location of the *egt* gene on the genome of HaSNPV. The *egt* gene is located over the junction of fragments *Hin*dIII-D and *Hin*dIII-I. The location of the polyhedrin gene (*ph*) is indicated (Chen *et al.*, 1997b). For the sequencing strategy the arrows indicate the direction of sequencing and the dots at the proximal end of the arrows indicate the use of specific primers. H = *Hin*dIII; RI = *Eco*RI; RII = *Eco*RII; RV = *Eco*RV; X = *Xba*I; S = *Sst*I. Bars indicate 1 kb and 100 bp, respectively. B. Nucleotide sequence of the HaSNPV *egt* gene and its flanking regions. The predicted amino acids are indicated by one-letter code designations below the nucleotide sequence. Putative transcription initiation and termination signals are bolded (see text). The putative signal sequence is in bold italics and possible N-linked glycosylation sequence (N-X-S/T) are underlined, the most conserved ones in bold. The *Hin*dIII (H), *Eco*RV (RV), *Eco*RII (RII) and *Xba*I (X) restriction enzyme sites in the coding sequence are shown

The 5'-noncoding region of the HaSNPV *egt* gene contained various control elements for early and late transcription. An early transcription initiation motif, CAGT, was observed at nt position -134 with respect to the ATG translational start. Putative enhancer elements, GATA and CGTGC, were observed at positions -174 and -30, respectively. Two TATA boxes were located at 308 and 100 nt upstream of the ATG translational start codon (Fig. 3-1b). The latter is located downstream of the CAGT motif and unlikely to be functional. A TATA box is one of the core elements of baculovirus early promoters (Guarino and Smith, 1992) and all baculovirus *egt* genes investigated so far contain at least one TATA box (Hu *et al.*, 1997). AcMNPV and *Lymantria dispar* (Ld) MNPV *egt* genes are transcribed from a site starting at 28 nt and 12 nt downstream of their TATA box, respectively (O'Reilly and Miller, 1990; Riegel *et al.*, 1994). A TAAG motif characteristic for baculovirus late transcription was found at position -92. At the 3' end of the HaSNPV *egt* ORF a polyadenylation signal, AATAAA, was found 30 nt downstream of the translation termination codon TAA. These observations suggest that the HaSNPV *egt* gene is likely to be transcribed early as well as late in infection, but transcription analysis should clarify this point.

The putative HaSNPV *egt* gene has an ORF of 1545 nt encoding a protein of 515 amino acids (aa) with a predicted size of 59.2 kilodalton (kDa) (Fig. 3-1b). The *egt* ORF has been compared to eleven other EGTs (Fig. 3-2; Table 3-1). The HaSNPV EGT contained six potential *N*-linked glycosylation consensus sequences (N-X-T/S), but it is not known if these sites are indeed glycosylated. Only two of these, at amino acid positions 67 and 295, are reasonably well conserved in most baculoviruses and possibly functional. *Egt* genes have now been identified from twelve baculoviruses: nine MNPVs, one GV and two SNPVs (see legend figure 2). A gross comparison of the HaSNPV *egt* gene with other baculovirus *egt* genes indicated a nucleotide homologues and amino acid sequence identity ranging from 45 to 59% and 42 to 53%, respectively (Table 3-1). The HaSNPV *egt* gene is most related to that of BusuNPV and SeMNPV with about 53% amino acid identity. It can be predicted from the similarity between HaSNPV and HzSNPV (Sun and Zhang, 1994) that their *egt* genes will be very similar.

EGT structure

Alignment of amino acid sequences shows that the baculovirus EGTs, including HaSNPV EGT, display similarity throughout their sequence (Fig. 3-2). Ten regions (I-X) with an identity of at least 45% can be discerned, well above the average overall identity of 35% among EGTs (Fig. 3-2) (Hu *et al.*, 1997). Regions V and VII are present in all animal, plant, nematode and bacterial UDP-glucosyltransferases (O'Reilly, 1995). Eight amino acids (HaSNPV EGT residues Ser33, Gly304, Trp353, Gln356, Gly373, Glu379, Pro386 and Gln396) are absolutely conserved in all known transferases and these are all but one (Ser33) located in either region V or region VII (Fig. 3-2). It has been suggested that these conserved amino acids may be involved in enzymatic activity and UDP-sugar binding of

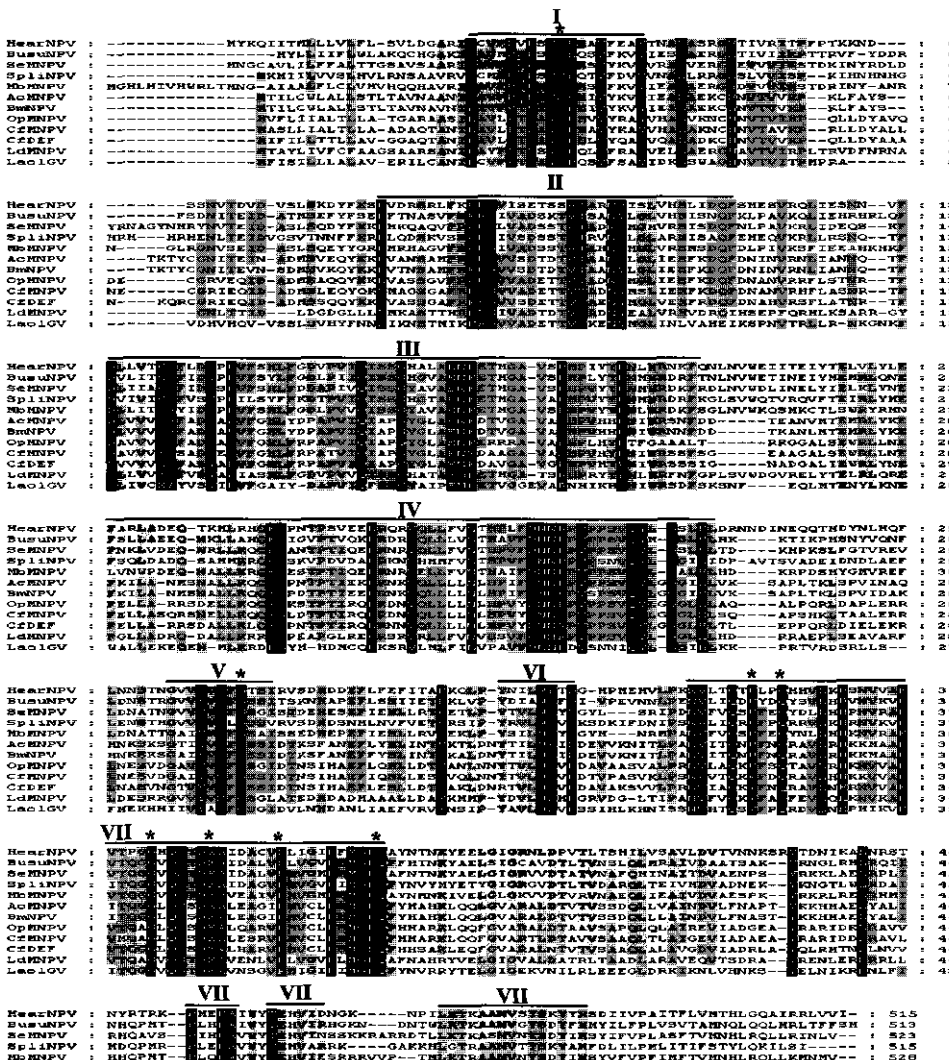


Figure 3-2. Alignment of the amino acid sequence of twelve baculovirus *egt* genes, from *Autographa californica* MNPV (AcMNPV, O'Reilly and Miller, 1989), *Bombyx mori* MNPV (BmNPV; Genbank L33180), two MNPVs from *Choristoneura fumiferana* (CfMNPV and CfDEF; Barrett *et al.* 1995), *Lymantria dispar* MNPV (Riegel *et al.*, 1994), *Spodoptera exigua* MNPV (SeMNPV; D. Zuidema, unpublished data), *Spodoptera littoralis* MNPV (SpliMNPV; Faktor *et al.*, 1995), *Orgyia pseudotsugata* MNPV (OpMNPV; Ahrens *et al.*, 1997), LaolGV (Smith and Goodale, 1997), *Mamestra brassicae* MNPV (MbMNPV; Clarke *et al.*, 1996) and *Buzura suppressaria* SNPV (BusuNPV; Hu *et al.*, 1997). GeneDoc software was used for homology shading. Two shading levels were set: black for 100% identity and grey for 70% identity. The conserved regions are indicated with numbers in Roman. The eight absolutely conserved amino acids among all the UDP-glucosyltransferases investigated are indicated by asterisk (*). The sources of the sequences and the abbreviations of the viruses are given in the text.

glucosyltransferases (O'Reilly, 1995). Regions VI, VIII, IX and X, and parts of regions I, III and IV are also conserved in the mammalian enzymes (O'Reilly, 1995; Hu *et al.*, 1997). The amino acid conservation in region II and part of III and IV is restricted to baculovirus EGTs, suggesting that these areas be involved for example in recognition of ecdysteroids in insects.

In the case of HaSNPV a putative signal sequence of 21 amino acids in length has been detected. As in all EGTs analyzed to date, HaSNPV EGTs has a hydrophobic N-terminus indicative of a putative signal peptide (Fig. 3-2). Among EGTs the length of the signal peptides is highly variable, but they all terminate in the highly conserved domain I of the mature EGT.

Table 3-1. Amino Acid sequence identity (%) of twelve baculovirus EGTs

	Busu NPV	SeM NPV	MbM NPV	Spli NPV	LdM NPV	AcM NPV	Bm NPV	Cf DEF	Laol GV	OpM NPV	CfM NPV
HaSNPV	**53.0**	**52.7**	**49.8**	**48.9**	**48.3**	**45.6**	**44.9**	**44.5**	**43.2**	**42.5**	**41.7**
BusuNPV		59.8	58.4	50.3	50.1	47.8	46.3	44.5	40.0	45.1	44.8
SeMNPV			71.0	48.4	54.0	50.3	49.2	48.0	42.3	49.6	48.7
MbMNPV				46.5	51.6	47.9	46.6	46.3	39.8	45.7	47.0
SpliNPV					46.8	41.6	40.1	40.7	45.0	41.0	39.8
LdMNPV						44.2	43.8	45.2	41.1	46.6	45.1
AcMNPV							95.5	64.0	39.7	59.5	62.7
BmNPV								63.3	40.4	58.5	61.7
CfDEF									37.1	74.8	74.1
LaolGV										46.6	37.4
OpMNPV											80.0

Phylogenetic analysis

The phylogeny of twelve EGTs was analyzed using the PAUP 3.1 computer program (Swofford, 1993). The most parsimonious tree using LaolGV as an outgroup is shown in figure 3-3. Bootstrap analysis (100 replicates) of the data showed that the tree topology is well supported. EGTs from AcMNPV, BmNPV, CfMNPV, OpMNPV and CfDEF, consistently belong to one clade. Within this clade two further subgroups, EGTs of AcMNPV and BmNPV on the one hand and CfMNPV, OpMNPV and CfDEF on the other, could be distinguished. EGTs of BusuNPV, MbMNPV and SeMNPV form another well-supported clade. Using LaolGV as an outgroup HaSNPV EGT forms a separate group suggesting that this EGT diverged considerably from the other NPVs at an early stage. A similar situation exists for SpliNPV (Fig. 3-3).

Our results on twelve EGTs accord with earlier phylogenetic trees based on seven (Barrett *et al.*, 1995), nine (Clarke *et al.*, 1996) and eleven (Hu *et al.*, 1997) *egt* genes, respectively. It remains to be seen whether the EGT phylogeny reflects the natural evolution of baculoviruses from a common ancestor gene. A better interpretation of this phylogeny can be obtained when more NPV and GV EGT sequences are available and a comparison can

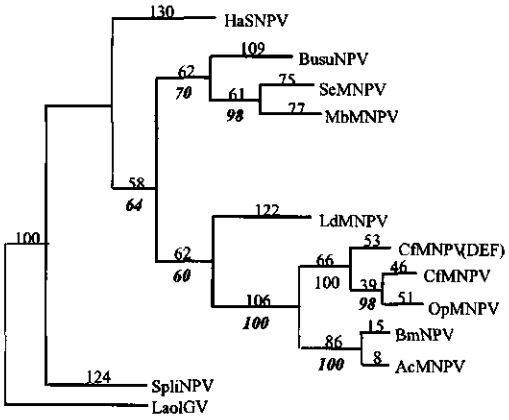


Figure 3-3. A phylogenetic tree of baculovirus *egt* genes. The most parsimonious tree for twelve baculovirus EGTs was constructed by using the 'branch-and-bound' algorithm of PAUP 3.1. LaolGV was chosen as an outgroup. Numbers above the lines indicate the phylogenetic distances. Numbers in bold italics indicate the frequency of a given cluster after bootstrap analysis (100 replicates).

be made with phylogenies based on other conserved baculovirus proteins (Zanotto *et al.*, 1993).

CONCLUDING REMARKS

The HaSNPV *egt* is the second *egt* gene discovered in SNPVs and the presence of this gene supports the view that all baculoviruses carry an *egt* gene. Comparison of the HaSNPV *egt* gene with other baculovirus *egt* genes confirms the general consensus motifs for glucosyltransferases as well as the specific motifs for EGTs. Transcriptional analysis, construction of deletion mutants and enzymatic analysis should indicate whether this *egt* gene is functional. HaSNPV is an effective biological agent for the control of the bollworm in cotton in China (Zhang, 1994). Deletion of its *egt* gene may enhance the efficacy of this virus. Information provided in this paper may help in the design of a proper strategy to achieve this goal.

ACKNOWLEDGEMENTS

We are grateful to Douwe Zuidema for reviewing this manuscript. This research was supported by a joint grant from the Chinese Academy of Science and the Royal Netherlands Academy of Sciences (contract 96CDP010) and a grant (no. 39670028) to Z. H. Hu from the National Natural Science Foundation of China. Z. H. Hu is a recipient of a PhD sandwich fellowship from the Wageningen Agricultural University. We thank Dr G. F. Rohrmann, I. R. L. Smith and D. Zuidema for sharing EGT sequence information of OpMNPV, SeMNPV and LaolGV, respectively, prior to publication. The sequence data reported in this paper have been deposited in GenBank under the accession number AF000009.

Chapter 4

IDENTIFICATION, SEQUENCE ANALYSIS AND PHYLOGENY OF THE *lef-2* GENE OF *Helicoverpa armigera* SINGLE-NUCLEOCAPSID BACULOVIRUS

SUMMARY

The baculovirus late expression factor 2 (LEF-2) is involved in DNA replication and most likely functions as a primase processivity factor. *Lef-2* genes have been found in MNPVs and in GVs but not yet in SNPV. Here, a *lef-2* gene homology was identified from SNPV of HaSNPV. The HaSNPV *lef-2* open reading frame (ORF) is 696 nt long, encoding a putative protein of 232 aa with a M_r of about 26 kDa. The 5'-noncoding region contains two early (CAGT) consensus motifs for transcription initiation and three TATA boxes. *Lef-2* transcripts started at C, 29 nt upstream of a putative translational start codon. A putative polyA signal, AATAAA, was found 76 nt downstream of the translation stop codon. The HaSNPV LEF-2 has a low but significant degree of similarity (30%) to the LEF-2s of fifteen other baculoviruses of which nine were newly determined. The N-terminal half of the LEF-2 proteins contains one (I) and the C-terminal half two (II and III) conserved domains. Sixteen amino acids are absolutely conserved in those LEF-2s investigated and are probably critical for LEF-2 function. A phylogenetic tree of sixteen baculovirus LEF-2 proteins was constructed and appeared to be comparable to a tree for ecdysteroid UDP-glucosyltransferases. The genomic location of the *lef-2* genes relative to polyhedrin/ granulin and the clade structure of the gene trees suggest that genome organization and gene phylogeny are useful parameters to study the evolutionary history of baculoviruses. These two independent approaches also give a more complete picture of the ancestral relationship among baculovirus.

This chapter has been published as

Chen, X., IJkel, W. F. J., Dominy, C., Zanotto, P. de A., Hashimoto, Y., Faktor, O., Hayakawa, T., Wang, C. H., Prekumar A., Mathavan, S., Krell, P. J., Hu, Z.and J. M. Vlak 1999, Identification, sequence analysis and phylogeny of the *lef-2* gene of *Helicoverpa armigera* single-nucleocapsid baculovirus. *Virus Research* 65: 21-32.

Chapter 4

INTRODUCTION

Baculovirus infection of insect larvae initiates a cascade of molecular and cellular processes ultimately leading to the death of the insect and the production of large amounts of polyhedral occlusion bodies containing rod shaped virions (Miller, 1997). DNA replication is a key process in baculovirus replication, in which *cis*- and *trans*-acting elements are involved (Kool *et al.*, 1994). The *cis*-acting elements encompass palindromic repeats in homologous regions (*hr*), which are dispersed in baculovirus genomes, and unique non-homologous regions (Lu *et al.*, 1997). In addition to DNA polymerase and DNA helicase, three late expression factors (LEFs) are involved as *trans*-acting elements in DNA replication. The *lef* genes involved most likely encode a primase (*lef-1*), a primase-processivity factor (*lef-2*) and a single-stranded DNA binding protein (*lef-3*).

In *Autographa californica* multi-nucleocapsid nucleopolyhedrovirus (AcMNPV) infected insect cells, LEF-2 is essential for DNA replication (Kool *et al.*, 1994; Lu and Miller, 1995) and late gene expression (Passarelli and Miller, 1993, Merrington *et al.*, 1996). In addition to AcMNPV, *lef-2* genes have also been found in *Orgyia pseudotsugata* MNPV (OpMNPV) (Ahrens and Rohrmann, 1995), *Bombyx mori* NPV (BmNPV) (Gomi *et al.*, 1999), *Lymantria dispar* NPV (LdMNPV) (Kuzio *et al.*, 1999), *Anagrapha falcifera* NPV (AnfaNPV) (Federici and Hice, 1997), *Rachiplusia ou* NPV (RaouNPV) (GenBank number AF68270) and *Cydia pomonella* granulovirus (CpGV) (Jehle *et al.*, 1997). Some partial *lef-2* sequences have been found in *Perina nuda* MNPV (PenuMNPV) (Chou *et al.*, 1996) and *Anticarsia gemmatalis* MNPV (AgMNPV) (Zanotto *et al.*, 1992). Thus, *lef-2* genes have been found in MNPVs and GVs, but have not yet been identified in single-nucleocapsid NPVs (SNPV). Baculoviruses have been taxonomically subdivided into two genera, NPV and GV (Murphy *et al.*, 1995). A further distinguishing feature among NPVs is the presence of single (S) and multiple (M) nucleocapsids per virion. This paper investigates whether there is phylogenetic support for this distinction.

Helicoverpa armigera SNPV (HaSNPV) is widely used as a bioinsecticide in Southeast Asia, notably in the People's Republic of China (Zhang, 1989). The virus is closely related to *H. zea* SNPV (Chen *et al.*, 1997a) and infects a range of *Helicoverpa* species. The major occlusion body protein (polyhedrin) and ecdysteroid UDP-glucosyltransferase (*egt*) genes of HaSNPV have been recently described (Chen *et al.*, 1997a; Chen *et al.*, 1997b) and a physical map of the 130 kilobase pairs-long viral DNA has been constructed (Chen *et al.*, 2000a). The analysis and location of genes allow a comparison of the genomic organization of baculoviruses as well as a study of baculovirus relatedness. In this paper we describe the characterization and sequence analysis of a putative *lef-2* gene from HaSNPV and compare its product LEF-2 with a number of baculoviruses LEF-2 proteins by sequence, phylogenetic and genomic analysis.

MATERIALS AND METHODS

Viruses

The HaSNPV was isolated from diseased *H. armigera* larvae found in the Hubei province of the People's Republic of China and produced in fourth instar larvae (Zhang, 1989). The polyhedra were purified by sucrose gradient centrifugation and stored at -20• until use. Viral DNA was isolated from polyhedra according to Sun and Zhang (1994) and digested with restriction enzymes according to the manufacturer's protocols (Life Technologies/BRL).

Cloning and sequencing

A HaSNPV *Hin*dIII library was constructed by shotgun cloning of *Hin*dIII-digested HaSNPV DNA into pTZ19R according to the procedures described by Sambrook *et al.* (1989). The termini of selected *Hin*dIII fragments were sequenced using universal primers. Further sequencing was done by primer walking using specific oligonucleotide primers. The sequences were screened against databases (GenBank/EMBL) and searched for gene and sequence homologues.

The sequence for *lef-2* has been newly determined from baculoviruses of *Amsacta albistriga* (AmalNPV / AaMNPV), *Anticarsia gemmatalis* (AgMNPV) (Zanotto *et al.*, 1992), *Buzura suppressaria* (BusuNPV/BsSNPV) (Hu *et al.* 1998), *Choristoneura fumiferana* (CfMNPV) (Barrett *et al.*, 1995), *H. armigera* (HaSNPV/HearSNPV) (Zhang, 1989), *Perina nuda* (PenuNPV/PnMNPV) (Chou *et al.*, 1996), *Plutella xylostella* GV (PxGV) (Kondo and Yamamoto, 1998), *Spodoptera exigua* MNPV (SeMNPV) (Smits *et al.*, 1988) and *Xestia c-nigrum* GV (XcGV) (Goto *et al.*, 1998), respectively. The *lef-2* sequences have been deposited in GenBank (see legend Fig. 4-2).

Transcriptional analysis

Total RNA was isolated from *H. zea* AM1 cells infected with hemolymph-derived HaSNPV-infected *H. armigera* larvae. Cells were infected at a multiplicity of infection of 10 $TCID_{50}$ units per cell and the total RNA was isolated using TRIzol (Gibco-BRL) at 24 hours post infection. The 5' end of the *lef-2* transcript was determined using a 5' RACE kit (Boehringer). First strand cDNA was synthesized with a *lef-2* specific primer VW239 (AAA TCG AAG GGT TGC ACA AC). After degradation of the mRNA template by the RNase H and tailing the cDNA with dATP, cDNA was amplified by PCR using the oligo dT-anchor primer and a nested *lef-2* primer VW232 (TGG CTC TCT CGA TAC ACT G). For the second PCR the oligo dT-anchor primer and primer VW232 were used. The 5' RACE product was cloned into pGEM-T and sequenced.

Computer analysis

Sequences were analyzed with the aid of the UWGCG computer program (Devereux *et*

al., 1984, release 10.0). DNA and deduced amino acid sequences were compared with the updated GenBank/EMBL database. The alignment of LEF-2 proteins using PileUp in the UWGCG was performed as the input to construct the most parsimonious trees with the branch-and-bound algorithm of PAUP 3.1 (Swofford, 1993). Bootstrap analysis (100 replicates) was performed by using branch-and-bound search algorithm. GeneDoc software was used for similarity shading and scoring among the aligned sequences.

RESULTS

The HaSNPV *lef-2* gene was identified after sequencing the ends of the 6.7 kb *Hin*dIII-I fragment (Fig. 4-1) and comparison of the sequences with databases (GenBank/EMBL). A detailed physical map of this fragment was constructed and the putative *lef-2* gene and flanking regions were sequenced from both strands using specific primers by the dideoxy-nucleotide chain termination method (Sanger *et al.*, 1977). The *lef-2* gene was located around map unit (m.u.) 17.5 of the HaSNPV genome (Chen *et al.*, 1999). In addition, a 201 nucleotides long ORF encoding a putative protein of 67 amino acids with homology to AcMNPV ORF111 (Ayres *et al.*, 1994) was found downstream of HaSNPV *lef-2*.

The putative *lef-2* open reading frame (ORF) is 696 nucleotides long encoding a protein of 232 amino acids with a predicted size of about 26 kDa (Fig. 4-2). The ORF size is considerably larger (60 nt) than the *lef-2* genes of AcMNPV (Passarelli and Miller, 1993), OpMNPV (Ahrens and Rohrmann, 1995), BmNPV (Sriram and Gopinathan, 1998), AnfaMNPV (GenBank U64986) and CpGV (Jehle *et al.*, 1997). The additional nucleotides were located at the 5' end of the ORF. Most notable are seven codons for contiguous threonine residues in this extension of HaSNPV LEF-2. The 5' noncoding region contains two early consensus transcription motifs CAGT located at –219 and –329 nucleotides, respectively, and three TATA boxes located at –91, -106 and –308 nucleotides, upstream of the translational start codon of *lef-2*. The actual 5' end of the *lef-2* transcript was determined by 5' RACE analysis of infected-cell RNA and found at the C at position -29 relative to the putative translational start. A polyadenylation signal,

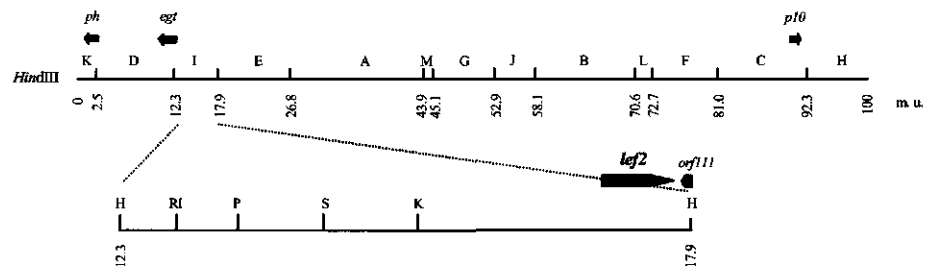


Figure 4-1. Location of the *lef-2* gene on the genome of HaSNPV. The *lef-2* gene is on fragment *Hin*dIII-I (m.u. 12.3 - 17.9). The location of the polyhedrin *(ph), p10* and *egt* genes is indicated (Chen *et al.*, 1997a; Chen *et al.*, 1997b). H = *Hin*dIII; K = *Kpn*I; P = *Pst*I; RI = *Eco*RI; S = *Sst*I.

AATAAA, is located 72 nucleotides downstream of the translation termination codon of *lef-2* (TGA). HaSNPV LEF-2 contains three potential N-linked glycosylation sites (N-X-S/T) and one protein kinase C phosphorylation site ([S/T]-X-[R/K]) (Fig. 4-2).

The amino acid sequence of the putative HaSNPV LEF-2 protein has been compared to those of fifteen other LEF-2 proteins (Fig. 4-2). In addition to AcMNPV, OpMNPV, BmNPV, LdMNPV, RaouNPV, AnfaNPV and CpGV, partial *lef-2* sequences were reported for AgMNPV (Zanotto *et al.*, 1992) and PenuNPV (Chou *et al.*, 1996). The sequencing of the latter two genes has now been completed. *Lef-2* genes have also been identified in CfMNPV, SeMNPV, PenuNPV, BusuNPV, AmalNPV, PxGV and XcGV and the complete sequences of these *lef-2* genes have also been obtained. A comparison of the *lef-2* genes indicated an amino acid identity ranging from more than 93% among AcMNPV, AnfaNPV, RaouNPV and BmNPV, 51-80% among AgMNPV, OpMNPV, CfMNPV, AmalNPV and PenuNPV to less than 40% among BusuNPV, HaSNPV,

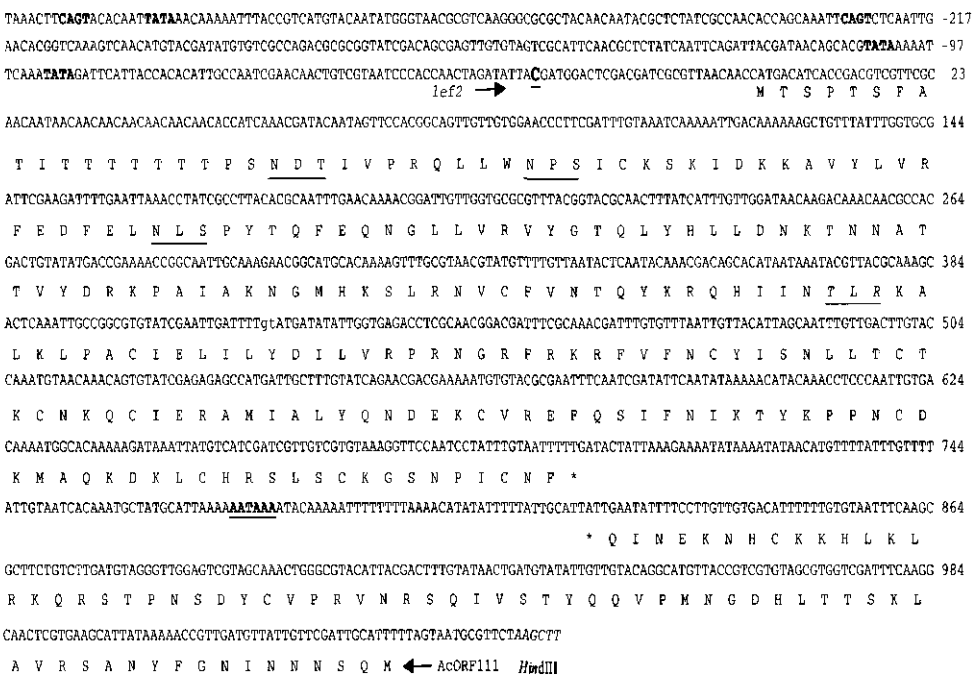


Figure 4-2. Nucleotide sequence of the HaSNPV *lef-2* gene and part of the downstream AcORF111 homologue. The deduced amino acids are indicated by one-letter code designations below the nucleotide sequence. Putative consensus baculovirus transcription initiation and termination signals are bolded (see text). The transcription initiation site is indicated (C in bold/underlined); the A in the putative ATG start condon is taken as +1. The putative N-linked glycosylation sites (N-X-S/T) are underlined and a possible protein kinase C phosphorylation site (T-L-R) is italicized and underlined. The *Hin*dIII (H) restriction enzyme site in the sequence is shown

Table 4-1. Amino acid identities of sixteen baculovirus LEF-2s

	Raou	Ac	Bm	Op	Penu	Cf	Ag	Amal	Ha	Se	Busu	Ld	Cp	Pl	Xc
Anfa	99	95	93	52	52	53	53	51	**31**	38	36	33	15	13	16
Raou		95	93	52	53	54	54	51	**31**	39	36	33	15	13	16
Ac			94	52	53	53	54	50	**32**	39	36	3	15	13	16
Bm				53	53	53	54	50	**32**	41	36	34	15	13	15
Op					97	80	68	42	**31**	34	30	34	16	12	15
Penu						80	68	43	**31**	35	31	35	16	12	15
Cf							66	46	**30**	33	31	34	16	11	13
Ag								42	**29**	33	31	32	14	12	14
Amal									**31**	36	32	28	16	14	15
Ha										**39**	**39**	**35**	**13**	**13**	**15**
Se											40	38	14	14	18
Busu												38	15	13	14
Ld													16	12	17
CP														24	34
Px															24

SeMNPV and LdMNPV. The overall amino acid identity with the three GV LEF-2s is very low (11-16%). Alignment shows the identity of HaSNPV with fifteen other baculoviruses ranging from 13% to 39% (Table 4-1).

Alignment of the LEF-2 indicated the presence of sixteen conserved amino acids probably critical in LEF-2 function. Most notable is the highly conserved Cys-rich C-terminal half of the protein. It is interesting to note that at the C-terminus of PxGV LEF2 is much longer while the CpGV (Jehle *et al.*, 1997) and XcGV LEF-2s are considerably shorter than the average LEF-2. Sixteen amino acids (HaSNPV LEF-2 residues Pro32, Tyr45, Thr60, Phe62, Gly66, Gly71, Leu76, Cys110, Arg147, Arg154, Cys167, Cys170, Cys174, Tyr173, Lys188 and Cys189) are absolutely conserved in all investigated LEF-2s. Three conserved regions (regions I-III) were identified. Region I was placed in the N-terminal half of the proteins with six conserved amino acids. Region II is a Cys-rich region which included a zinc finger-like domain, C-x(2)-C-x(3)-C-x(8)-H-x(2)-H, matching the C2H2 type's consensus pattern C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H. However, the zinc finger domain was only found in LEF-2s of AcMNPV, AnfaNPV and RaouNPV. The other LEF-2s lacked one or two of the conserved histidines. Three other conserved amino acids (Tyr174, Lys188 and Cys189) were present. Domain III was only found in nucleopolyhedroviruses and is highly conserved.

The availability of sixteen *lef-2* genes allowed the determination of a phylogenetic relationship among LEF-2 proteins (Fig. 4-4). An unrooted parsimonious tree was

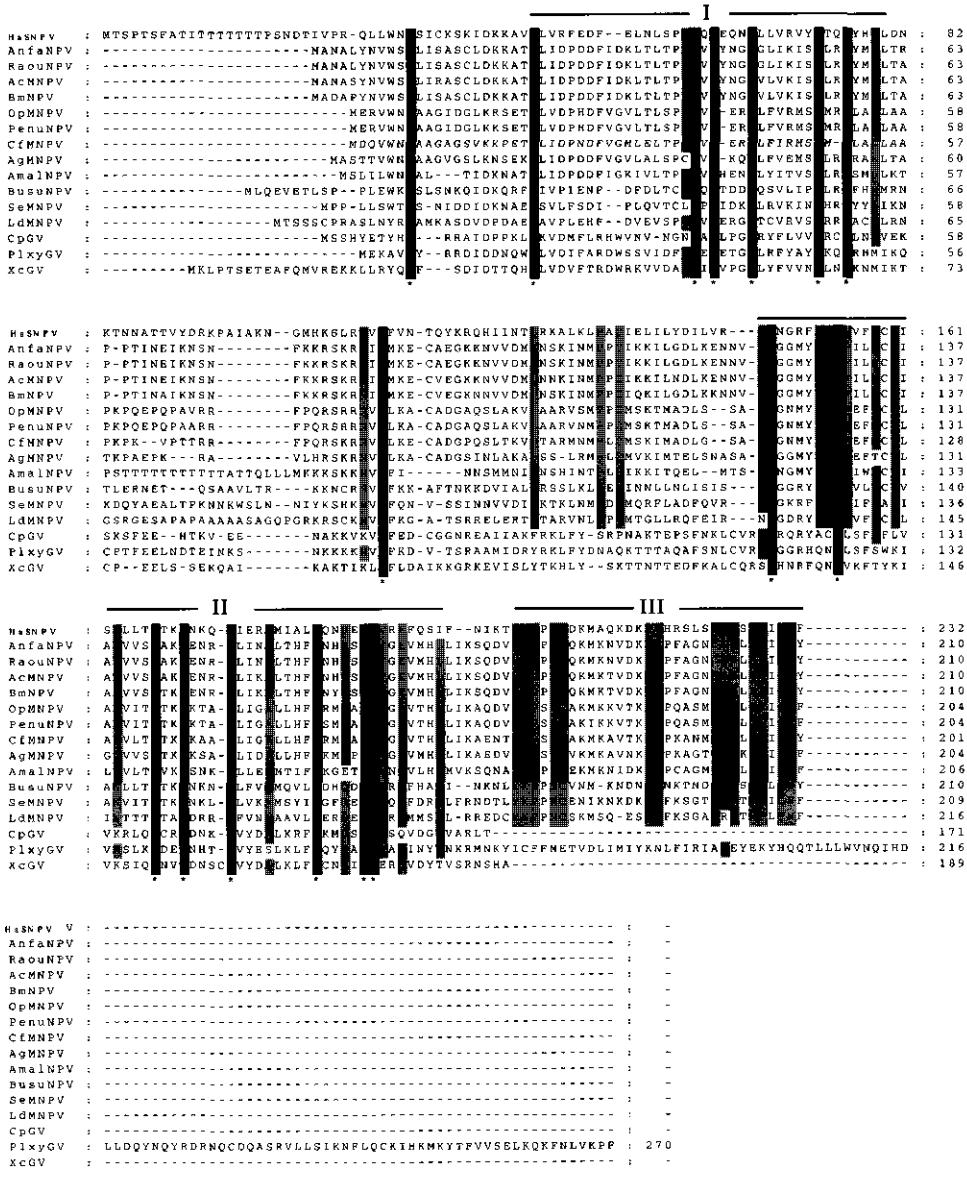


Figure 4-3. Alignment of the amino acid sequences of 16 baculovirus LEF-2 proteins, from AcMNPV (Passarelli and Miller, 1993), AgMNPV (GenBank AF168706), AmalNPV (GenBank AF118112), AnfaNPV (Federici and Hice, 1997), BmNPV (Gomi et al., 1999), BusuNPV (GenBank AF060564), CfMNPV (GenBank AF071497), CpGV (Jehle et al., 1997), HaSNPV (GenBank AF136502), LdMNPV (Kuzio et al., 1999), RaouNPV (GenBank AF068270), OpMNPV (Ahrens et al., 1997), PenuNPV (GenBank AF162787), PlxyGV (GenBank AB029255), SeMNPV (GenBank AF169824), and XcGV (GenBank AF162221). GeneDoc software was used for homology shading. Two shading levels were set: black for 100% identity and grey for 80% identity. The conserved regions are indicated with Roman numerals (I-III), while the conserved amino acid residues are denoted by an asterisk (*).

calculated with the PAUP branch-and-bound search algorithm followed by bootstrap analysis to assess the variability of the phylogeny produced. On the basis of this analysis AcMNPV, BmNPV, RaouNPV and AnfaMNPV clustered together, as was the case for OpMNPV, CfMNPV, PenuMNPV and AgMNPV with high bootstrap values. These two clusters and AmalNPV represent group I baculoviruses (Zanotto *et al.*, 1993; Cowan *et al.*, 1994). The two SNPVs, HaSNPV and BusuNPV LEF-2s, also formed a cluster but this supported only by a low bootstrap value. These viruses, together with LdMNPV and SeMNPV, represent group II. The GV group, including CpGV, XcGV and PxGV, formed a highly distinct cluster supported by high bootstrap values and separated from the NPV clusters. The LEF-2 tree is very similar in shape to the EGT tree (Chen *et al.*, 1997b) separating the respective baculoviruses in similar clusters.

Not only the gene homology but also the location of *lef-2* may give information on the relatedness among baculoviruses. In the group I baculoviruses AcMNPV, AnfaMNPV and BmNPV, the *lef-2* gene is located proximal to the polyhedrin gene (Fig. 4-5). This also appeared to be the case for AgMNPV (Zanotto *et al.*, 1992), PenuNPV (Chou *et al.*, 1996) and CfMNPV (this paper). These viruses represent group I. In the case of PxGV, CpGV and XcGV the *lef-2* genes were located about 77 kb, 90 kb and 149 kb away (clockwise) from the granulin locus (Jehle *et al.*, 1997). In HaSNPV and BusuNPV the

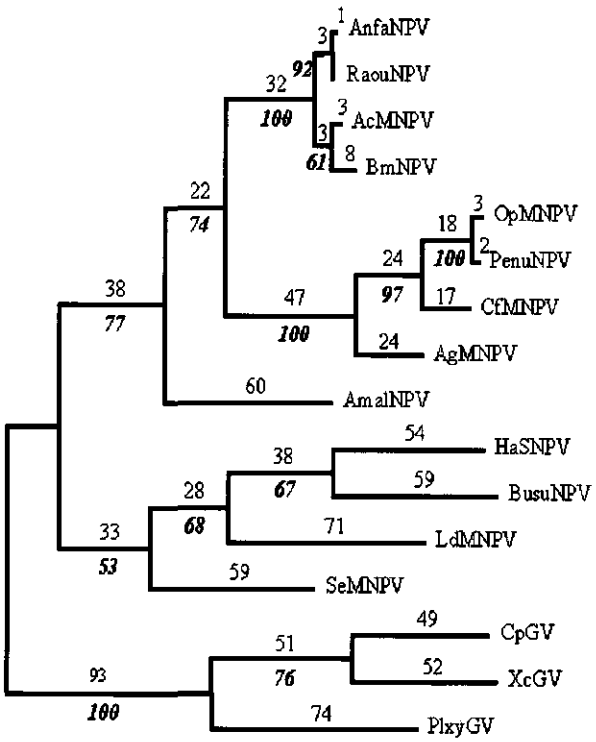


Figure 4-4. A phylogenetic tree of baculovirus LEF-2 sequences. The amino acid sequences of sixteen baculovirus *lef-2* genes were used. The most parsimonious tree for LEF-2 was constructed by using the 'branch-and-bound' algorithm of PAUP3.1. Numbers above the lines indicate the phylogenetic distances. Numbers in bold italics indicate the frequency of a given cluster after bootstrap analysis (100 replicates).

lef-2 genes are present 20 kb (counter-clockwise) and 25 kb (clockwise) away from polyhedrin. Also in LdMNPV the *lef-2* gene is found about 26 kb downstream from polyhedrin. SeMNPV *lef-2* is 120 kb downstream from polyhedrin, but 15 kb if the circular map is mirrored.

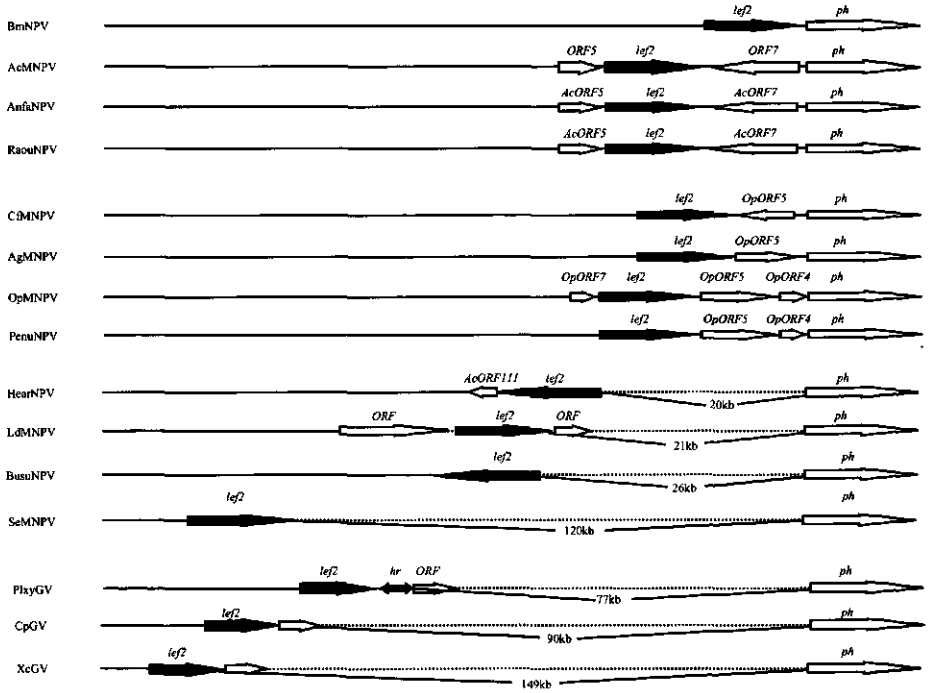


Figure 4-5. Genomic location of baculovirus *lef-2* genes relative to polyhedrin (*ph*). Other ORFs are either unnamed (no homology found) or named according to AcMNPV or OpMNPV homologies described by Ayres *et al.* (1994) and Ahrens *et al.* (1997), respectively. Baculovirus abbreviations were used as described in Materials and Methods and in the legend of figure 2.

DISCUSSION

Lef-2 is a baculovirus gene, of which the product LEF-2 has been implied to play a key role in viral DNA replication (Kool *et al.*, 1994; Lu and Miller, 1995) as well as in transcription of very late baculovirus genes (Passarelli *et al.*, 1993, Merrington *et al.*, 1996, Sriram and Gopinathan, 1998). LEF-2 interacts with the *lef-1* gene product (LEF-1) as evidenced from the "yeast two-hybrid" analysis system (Evans *et al.*, 1997). Sequence analysis and comparison with databanks suggested that LEF-1 is a putative primase and that its interaction with LEF-2 is compatible with the putative function of LEF-2 as a primase processivity factor. However, direct experimental evidence for this supposition is lacking.

The HaSNPV *lef-2* gene has been identified on a 6.7 kb *Hin*dIII fragment as an ORF with

multiple putative consensus transcription initiation and termination sites (Fig. 4-1b). Transcription analysis indicated that a much more downstream start site (C at -29) in a hitherto unknown sequence motif TACG was used. In BmNPV the 5' untranslated leader of *lef-2* appeared to be more than 300 nucleotides long (Sriram and Gopinathan, 1998). It is also clear that the second putative methionine is used as a translation initiation condon since the first in-frame methionine at -27 is located at the transcriptional start site. N-terminal analysis of HaSNPV LEF-2 should confirm this hypothesis.

Three regions (I-III) with conserved sequence elements have been identified after alignment of sixteen LEF-2 proteins (Figure 4-2). Region I is located in the N-terminal portion of LEF-2 from amino acid residues 45 to 80 (in HaSNPV LEF-2), and is marked by six absolutely conserved amino acids. Hydrophobicity analysis did not reveal further structural homology in this domain or obvious analogies with other proteins such as herpesvirus replication proteins (data not shown). In AcMNPV the region from amino acid residue 20 to 60 appeared to be involved in the interaction with LEF-1 as evidenced from deletion mutagenesis of *lef-2* and hence the lack of interaction with LEF-1 in the "yeast two-hybrid" system (Evans *et al.*, 1997). The divergence in this region among different baculoviruses may suggest that the interaction between LEF-1 and LEF-2 is virus species specific. This view is supported experimentally by plasmid-dependent DNA replication assays, where a heterologous *lef-2* gene (SeMNPV) was unable to replace a homologous *lef-2* gene (AcMNPV) to support transient baculovirus DNA replication (Broer *et al.*, 1998).

Region II is the most conserved region of LEF-2. Internal deletions of this region abolished the transcriptional activity of *lef-2* (Sriram and Gopinathan, 1998). In this domain a typical zinc finger motif (C2H2 type), which was first found in the *Xenopus laevis* transcription factor IIIA (Evans and Hollenberg, 1988) and has since been found to be ubiquitous in transcription factors and other DNA-binding protein structures, was found only in LEF-2s of AcMNPV, RaouMNP and AnfaNPV. The other baculoviruses lack one or two conserved histidines casting some doubts about the role of a zinc finger in LEF-2 function in general. An asparagine residue (D178) has been identified in AcMNPV LEF-2 that on mutation inactivates the late gene activation function of LEF-2 without affecting viral DNA replication (Merrington *et al.*, 1996). The asparagine, however, is not conserved in all baculovirus LEF-2s investigated (Fig. 4-2). A putative function of domain III, which is found only in NPVs and not in GVs, is still unknown. This domain is highly conserved and may be involved in specific steps of NPV transcription.

The pivotal role of LEF-2 in baculovirus replication and transcription suggests that this factor is important in the baculovirus life cycle and that it is to the advantage of the virus to maintain this gene during evolution. Sixteen baculoviruses, representing MNPVs, SNPVs and GVs appeared to contain a *lef-2* gene and it is most likely that all

baculoviruses have such a gene. Phylogenetic analysis indicates that the *lef-2* genes of NPVs have a common ancestor (Fig. 4-3). When compared to EGT (Chen *et al.* 1997a), a similar tree structure is observed for LEF-2. This may suggest that *egt* and *lef-2* genes have a common evolutionary history. This is a notable observation since the evolutionary pressure on these two genes is quit different. LEF-2 may have evolved along with other factors involved in DNA replication, known to be a very conserved process (Kornberg and Baker, 1992). In contrast, EGT is an auxiliary protein, not essential for virus replication, but optimized for abrogation of molting of the insect host (O'Reilly, 1997). Most likely *egt* genes have been picked up from the ancestral host at an early stage of baculovirus-insect host interaction along with the other genes common to baculoviruses such as *lef-2*.

The evolution of baculovirus genes is constrained by their function. The location of genes on the genome, however, might not and may therefore be another informative way to describe their evolutionary history. In baculoviruses a high percentage of genes has been maintained and their genomic location may be a reflection of that history (Hu *et al.*, 1998). On the basis of gene phylogeny using polyhedrin, LEF-2 and EGT (Zanotto *et al.*, 1993, Cowan *et al.*, 1994, Chen *et al*, 1997b) the location of *lef-2* genes is proximal to *polyhedrin* in the group I baculoviruses, whereas in the other viruses this is not the case (Fig. 4-4). This is in line with the clustering according to the phylogenetic analysis of LEF-2 and EGT. The data presented here further suggest that the genomic organization and gene phylogeny reflect a similar evolutionary history. Hence both parameters are useful for constructing baculovirus phylogenies (Hu *et al.*, 1998).

ACKNOWLEDGEMENTS

We are grateful to Dr. M. Erlandson for critical review of the manuscript. This research was supported by a joint grant from the Chinese Academy of Science and the Royal Netherlands Academy of Sciences (contract 97CDP010) and a grant (no. 39670028) to Z.H. Hu from the National Natural Science Foundation of China. X. Chen is a recipient of a PhD Sandwich Fellowship from the Wageningen University.

Chapter 5

GENOMIC ORGANIZATION OF *Helicoverpa armigera* SINGLE-NUCLEOCAPSID NUCLEOPOLYHEDROVIRUS

SUMMARY

The genome organization of the *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus (HaSNPV) was investigated and compared to that of other baculoviruses. A detailed physical map was constructed for the restriction enzymes *Bam*HI, *Bgl*II, *Eco*RI, *Hin*dIII, *Kpn*I, *Pst*I, *Sac*I and *Xho*I. The 130.1-kilobase pairs (kb) long viral genome was cloned as restriction fragments into a plasmid library, which covered 97.5% of the viral genome. From this plasmid library about 45 kb of dispersed sequence information was generated. Fifty-three putative open reading frames (ORFs) with homology to ORFs of other baculoviruses were identified and their location on the genome of HaSNPV was determined. The arrangement of the ORFs along the HaSNPV genome is significantly different from the multiple-nucleocapsid NPVs, AcMNPV, BmNPV and OpMNPV, which have a highly collinear genome, and the granulovirus XcGV. In contrast the genomes of HaSNPV and SeMNPV are highly collinear. This close relatedness between an MNPV and an SNPV is supported by the phylogeny of specific genes of these two viruses and suggests that the morphotype (S or M) is only a taxonomic and not a phylogenetic denominator. The data presented here also give further support to the view that the gene distribution along baculovirus genomes can be used as a character to describe baculovirus relatedness independent of gene phylogeny. Five homologous regions (*hr*s) were located on the genome of HaSNPV. This is also the first report of the existence of *hr*s in SNPVs and this observation suggests that *hr*s are a common feature of baculovirus genomes.

This chapter has been published as

Chen, X., Li, M., Sun, X., Hu, Z. and J. M. Vlak. 2000, Genomic organization of *Helicoverpa armigera* single–nucleocapsid nucleopolyhedrovirus. *Archives of Virology* 145:2539-2555.

Chapter 5

INTRODUCTION

The Baculoviridae, a diverse family of more than 600 viruses, encompass two genera, the *Nucleopolyhedroviruses* (NPV) and the *Granuloviruses* (GV) (Murphy *et al.*, 1995). The virions are rod-shaped and contain a circular, double stranded DNA genome ranging from 80 to 160 kilobasepairs (kb) in size. The family is characterized by the occlusion of virions into large proteinaceous capsules or occlusion bodies (OBs). In NPVs the DNA-containing nucleocapsids are enveloped singly (S) or in multiples (M) into virions and this property is an important taxonomic denominator. Developments in molecular genetics and bioinformatics allow the study of the relatedness of baculoviruses as well as of their ancestral relationships by gene and genomic phylogeny.

Baculoviruses are pathogenic for arthropods, particularly insects of the order Lepidoptera, and cause a disease, which is often lethal. Because of their specificity, baculoviruses offer an attractive means to specifically control pest insects with minimal adverse environmental effects (Ignoffo, 1981; Moscardi, 1999). As compared with chemical insecticides, the major drawback of baculoviruses is their slow speed of action, since delay between virus application and feeding arrest can still result in significant crop damage. With genetic engineering techniques, the insecticidal properties of baculoviruses, in particular the reduced time to feeding arrest, have been enhanced by either deletion mutagenesis or by inserting insect-selective neurotoxin genes (Miller, 1995).

The sequence of six baculoviruses genomes, *Autographa californica* MNPV (AcMNPV) (Ayres *et al.*, 1994), *Bombyx mori* NPV (BmNPV) (Gomi *et al.*, 1999), *Orgyia psudotsugata* MNPV (OpMNPV) (Ahrens *et al.* 1997), *Lymantria dispar* MNPV (LdMNPV) (Kuzio, *et al.*, 1999), *Spodoptera exigua* MNPV (SeMNPV) (IJkel *et al.*, 1999) and *Xestia c-nigrum* GV (XcGV) (Hayakawa *et al.*, 1999) has been obtained and others are forthcoming. Although there are small inversions and insertions (or deletions), basically AcMNPV, BmNPV and OpMNPV have a similar gene content and genomic organization. According to the phylogenetic trees based on the polyhedrin, ecdysteroid UDP-glucosyltransferase (*egt*), DNA polymerase and late-essential-factor-2 (*lef-2*) genes these baculoviruses are members of Group I NPVs (Bulach *et al.*, 1999; Chen *et al.*, 1999; Cowan *et al.*, 1994; Hu *et al.*, 1997; Zanotto *et al.*, 1993). On the basis of these characters LdMNPV and SeMNPV are Group II NPVs and their gene content and overall genome organization indeed differs considerably from Group I NPVs (IJkel *et al.*, 1999; Kuzio *et al.*, 1999). However, Group II NPVs is still a somewhat heterogeneous collection of baculoviruses (SNPVs and MNPVs); the differences among its entirely sequenced members (LdMNPV and SeMMPV) are about as significant as when each member is compared to Group I baculoviruses. The morphotype (MNPV, SNPV) is a useful taxonomic denominator, but does not seem to reflect relatedness based on gene phylogeny (Chen *et al.*, 1999). Partial sequence information derived from the *Buzura suppressaria*

SNPV genome (BusuNPV) indicated that this SNPV has a distinct genomic organization as compared to AcMNPV, OpMNPV and BmNPV (Hu *et al.*, 1998). Analysis of the newly characterized LdMNPV and SeMNPV seems to support this view for this virus (IJkel *et al.*, 1999; Kuzio *et al.*, 1999). It has been postulated that the genomic organization can be an important character to determine the relatedness between baculoviruses (Hu *et al.*, 1998). A comparison with additional SNPVs could shed further light on the generality of this hypothesis.

A number of NPVs have been isolated world wide from insect species belonging to the genus *Helicoverpa* (Lepidoptera: Noctuidae), which include agricultural pests such as *Helicoverpa zea, H. virescens, H. armigera* and *H. punctigera* (Gettig and McCarthy, 1982). These *Helicoverpa spp.* NPVs can be divided morphologically into two groups, either MNPVs or SNPVs. The bollworm *H. armigera* is a major pest in China. A SNPV (HaSNPV) has been isolated and successfully used to control *H. armigera* in about 100,000 ha of cotton in China (Zhang, 1989; Zhang, 1994). HaSNPV is closely related to HzSNPV (Chen *et al.*, 1997a). A genetic approach can be adopted to improve the insecticidal properties of HaSNPV. This, however, requires insight in the genomic organization and gene regulation of this baculovirus.

In this paper we present a study characterizing genome organization of the HaSNPV (WIV-1 isolate, 4 strain). A detailed physical map of the viral genome was constructed and a genomic library was made by cloning the viral DNA as restriction fragments into plasmid vectors. The partial or total nucleotide sequence of these inserts was determined and used to identify potential open reading frames (ORFs) and baculovirus gene homologues by comparison with databases. A genetic map describing the order of fifty-three putative ORFs in the genome of HaSNPV was generated and compared to that of AcMNPV, SeMNPV, XcGV and BusuNPV by using GeneParityPlot analysis (Hu *et al.*, 1998). We also present the discovery, location, and partial characterization of homologous repeats (*hr*) regions in the HaSNPV genome.

METHODS

Virus and DNA

HaSNPV was isolated from diseased *H. armigera* larvae found in Hubei Province of China (Zhang, 1989) and purified by in vivo cloning (Sun *et al.*, 1998). The major genotype (G4) was selected for further analysis. The virus was produced in fourth instar larvae and OBs were purified by differential and rate zonal centrifugation (King and Possee, 1992) The DNA was isolated directly from purified OBs by incubation in an alkaline dissolution buffer (0.1 M Na_2CO_3, 0.01 M EDTA and 0.17 M NaCl), followed by proteinase-K and SDS treatment, phenol/chloroform extraction and dialysis (King and Possee, 1992). The purity of the DNA was determined spectrophotometrically.

Restriction endonuclease (REN) analysis and cloning of viral DNA fragments in plasmid vector

Viral DNA was digested with restriction enzymes *Bam*HI, *Bgl*II, *Eco*RI, *Hin*dIII, *Kpn*I, *Pst*I, *Sac*I and *Xho*I (GIBICO/BRL) and the fragments were separated by electrophoresis in 0.6% agarose gels at 45 V (1.5V/ cm) for 16 h. Lambda DNA digested with *Bam*HI-*Eco*RI- *Hin*dIII was used as molecular size standards. The presence of low molecular weight fragments was checked by Southern hybridization (Southern, 1975)

Viral DNA restriction fragments were cloned into the plasmid vector, pTZ19R, using standard techniques (Sambrook, 1989). Firstly, the shotgun method was applied to generate a library of clones. Restriction fragments, absent from the shotgun library, were isolated from agarose gels, purified by the GlassMilk method (GIBICO/BRL) and cloned individually into pTZ19R (Sambrook, 1989).

***In vitro* labeling of DNA and Southern blot hybridization.**

Virus DNA digested by various restriction enzymes was separated in 0.6% agarose gels and transferred to nitrocellulose filters. Plasmids containing cloned viral DNA fragments were labeled with α-^{32}P dATP using a nick translation kit (GIBCO-BRL) and used as probes in Southern blot hybridization (Southern, 1975) to detect overlapping fragments.

Construction of the physical map of the HaSNPV genome

The relative order of the restriction fragments on the viral genome was derived from the data of Southern blot hybridization and restriction enzyme mapping of cloned viral DNA fragments. Sequence data of the termini of cloned fragments were also used to establish and confirm the detailed physical map. PCR cloning and sequencing was adopted to confirm fragment junctions and to close gaps.

DNA sequencing and computer analysis

Plasmid DNA for sequencing was purified via Qiagen columns (Qiagen, Inc.). Partial sequencing was conducted from both ends of the cloned fragments using the universal forward and reverse pTZ19R primers. Selected regions of the genome were completely sequenced on both strands using either a series of overlapping clones containing nested deletions or using the "primer walking" method with custom-synthesized primers. Automatic sequencing was carried out at the Sequencing Facility in the Department of Molecular Biology of the Wageningen University and at the core facility of Queens University (Kingston, Ontario, Canada). The generated sequences were analyzed with UWGCG computer programs (release 10.0). The DNA and the deduced amino acids sequences were compared with the updated GenBank/EMBL, SWISSPORT and PIR databases using FASTA and BLAST. The nucleotide sequences of the complete ORFs of HaSNPV have been deposited in Genbank and were assigned the following accession numbers: U97657 (polyhedrin) (Chen *et al.*, 1997a) AF000009 (EGT) (Chen *et al.*,

1997b) AF136502 (LEF2) (Chen *et al.*, 1999) and AF197913 (P6.9) (Wang *et al.*, 2001). Sequences of other ORFs or fragments are available upon request.

Gene order in baculovirus genome

The genomic organization between HaSNPV on the one hand and AcMNPV, SeMNPV and XcGV on the other was studied by pairwise analysis of their genomes using GeneParityPlot (Hu *et al.*, 1998). Only those gene homologues identified so far in each pair of genomes were selected for analysis. In brief, the selected ORFs were renumbered starting with the polyhedrin gene as # 1 and ordering the other selected ORFs sequentially from left to right according to their relative occurrence on the respective genomes. The genomes were oriented with respect to each other in such a way that maximum parallel alignment of the highly conserved central region (Heldens *et al.*, 1998) was obtained. With the gene order of HaSNPV on the X-axis and the order of the gene homologues in the other viruses on the Y-axis, a correlation chart was plotted from which the genomic relatedness between two baculoviruses could be easily derived.

RESULTS AND DISCUSSION

Restriction endonuclease analysis of HaSNPV DNA

Digestion of HaSNPV DNA with *Bam*HI, *Bgl*II, *Eco*RI, *Hin*dIII, *Kpn*I, *Pst*I, *Sac*I and *Xho*I resulted in 11, 13, 25, 13, 6, 7, 6 and 9 fragments, respectively. The fragments were either identified by ethidium bromide staining (Fig. 5-1) or by Southern blot hybridization for the smaller fragments (data not shown). The fragments were designated alphabetically starting with A for the largest fragment for each enzyme digest as proposed by Vlak & Smith (1982). Restriction profiles of these digestions of HaSNPV DNA are shown in Fig. 5-1 and the approximate size of the fragments is summarized in Table 5-1. These sizes were estimated from the original agarose gel and adjusted according to the data obtained from the mapping and sequencing of the HaSNPV clones.

The total size of the HaSNPV is determined to be approximating 130.1 kb. This is in the same order of magnitude as the genomes of AcMNPV (133.9 kb) (Ayres *et al.*, 1994), BmNPV (128.4 kb) (Gomi *et al.*, 1999), OpMNPV (131.99 kb) (Ahrens *et al.*, 1997), and SeMNPV (135.6 kb) (IJkel *et al.*, 1999) but considerably smaller than LdMNPV (161 kb) (Kuzio *et al.*, 1999). The latter genome contains 16 so-called 'baculovirus repeated ORFs' or 'bro' genes accounting for most of the size difference. The estimation of the overall genome size of HaSNPV DNA is slightly higher than the estimates of the closely related HzSNPV DNA (119 kb or 125 kb) (Corsaro and Fraser, 1987; Knell and Summers, 1984), but this may be due to underestimating the sizes of the larger restriction fragments.

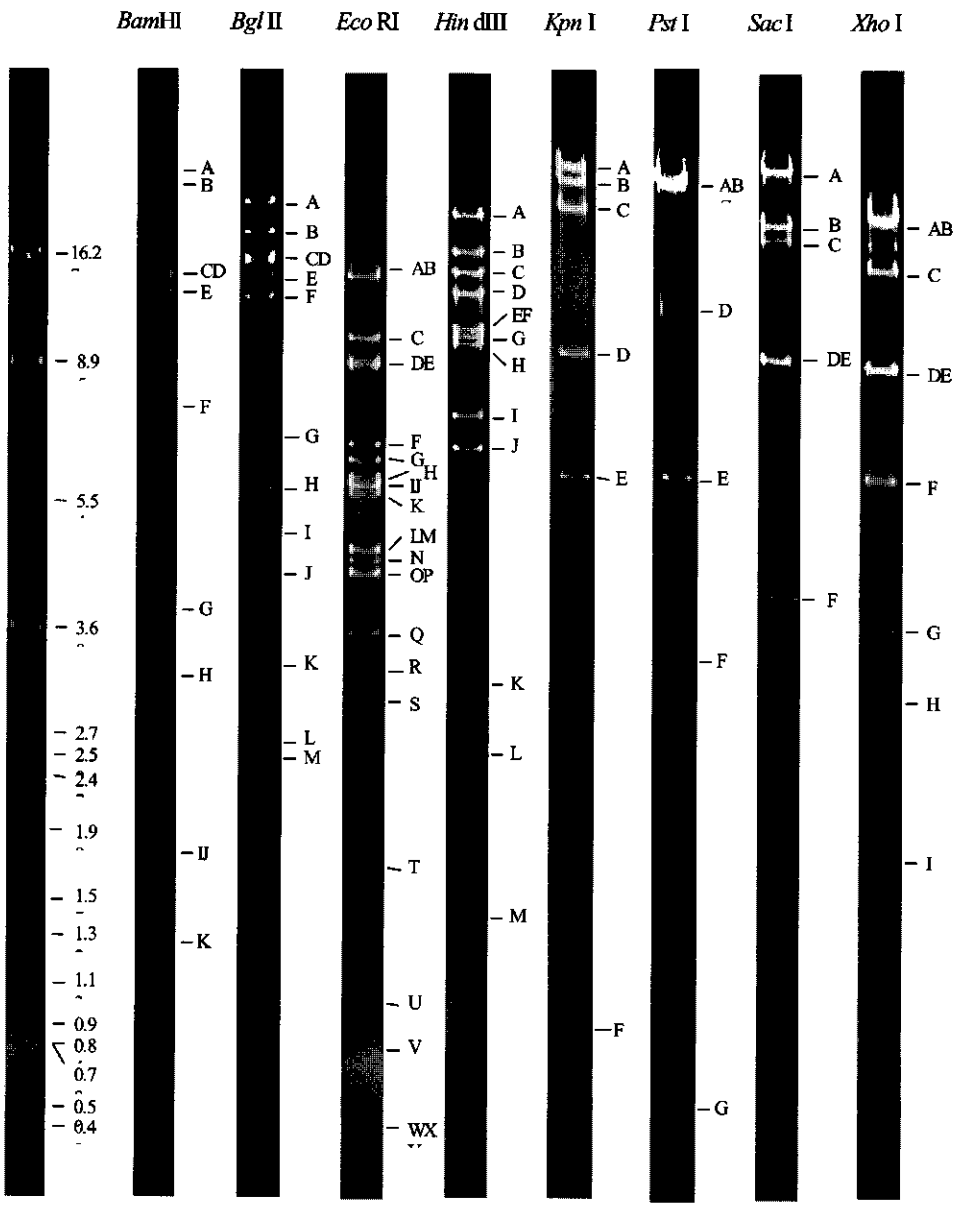


Figure 5-1. HaSNPV DNA digested with *Bam*HI, *Bgl*II, *Eco*RI, *Hin*dIII, *Kpn*I, *Pst*I, *Sac*I and *Xho*I and electrophoretically separated in 0.6% agarose. The fragments are named alphabetically, starting with A for the largest fragment. Lambda DNA digested with *Bam*HI-*Eco*RI-*Hin*dIII was used as molecular size markers with the sizes indicated in kilobase pairs.

Six of the 11 *Bam*HI fragments, 7 of the 13 *Bgl*II fragments, 16 of the 25 *Eco*RI fragments and 10 of 13 *Hin*dIII fragments were cloned into plasmid vector pTZ19R (bolded in Table 5-1). These fragments represent about 97.5% of the HaSNPV genome (Fig. 5-2) and provided sufficient information on the gene content

The restriction map of HaSNPV DNA

Data from Southern blot hybridization using HaSNPV fragment clones as probes and restriction enzyme-digested viral DNA immobilized on membrane blots were used to construct a linearized physical map of HaSNPV (Fig. 5-2). Data generated with probes of *Hin*dIII fragments were used to align these fragments. For regions of the genome in which *Hin*dIII clones were not sufficient to define the fragment order, cloned fragments from

Table 5-1. Size of restriction endonuclease fragments (kbp) of HaSNPV DNA (numbers in bold were cloned)

Fragment	*Bam*HI	*Bgl*II	*Eco*RI	*Hin*dIII	*Kpn*I	*Pst*I	*Sac*I	*Xho*I
A	37.3	24.5	**14.1**	22.2	55.5	39.0	65.0	36.5
B	31.8	18.5	**13.9**	16.5	34.2	36.8	22.3	34.6
C	**14.4**	**15.8**	**9.8**	14.7	23.6	32.3	19.3	20.0
D	14.0	**14.8**	**9.1**	**12.8**	9.8	11.8	9.7	11.0
E	**12.7**	13.7	**9.0**	**11.6**	6.1	6.1	9.4	10.9
F	**7.7**	12.1	**6.8**	**10.8**	0.9	3.4	4.4	7.0
G	3.9	**7.1**	**6.4**	**10.2**		0.6		4.4
H	**3.3**	5.9	6.0	**10.1**				3.5
I	**1.9**	**4.9**	**6.0**	**7.3**				2.2
J	**1.8**	**4.3**	**5.8**	**6.5**				
K	1.3	**3.4**	**5.6**	**3.2**				
L		**2.6**	**4.7**	**2.7**				
M		2.5	4.6	**1.5**				
N			**4.5**					
O			4.4					
P			4.3					
Q			**3.7**					
R			3.3					
S			3.1					
T			**1.7**					
U			**1.0**					
V			**0.8**					
W			0.5					
X			0.5					
Y			0.5					
Total				130.1				

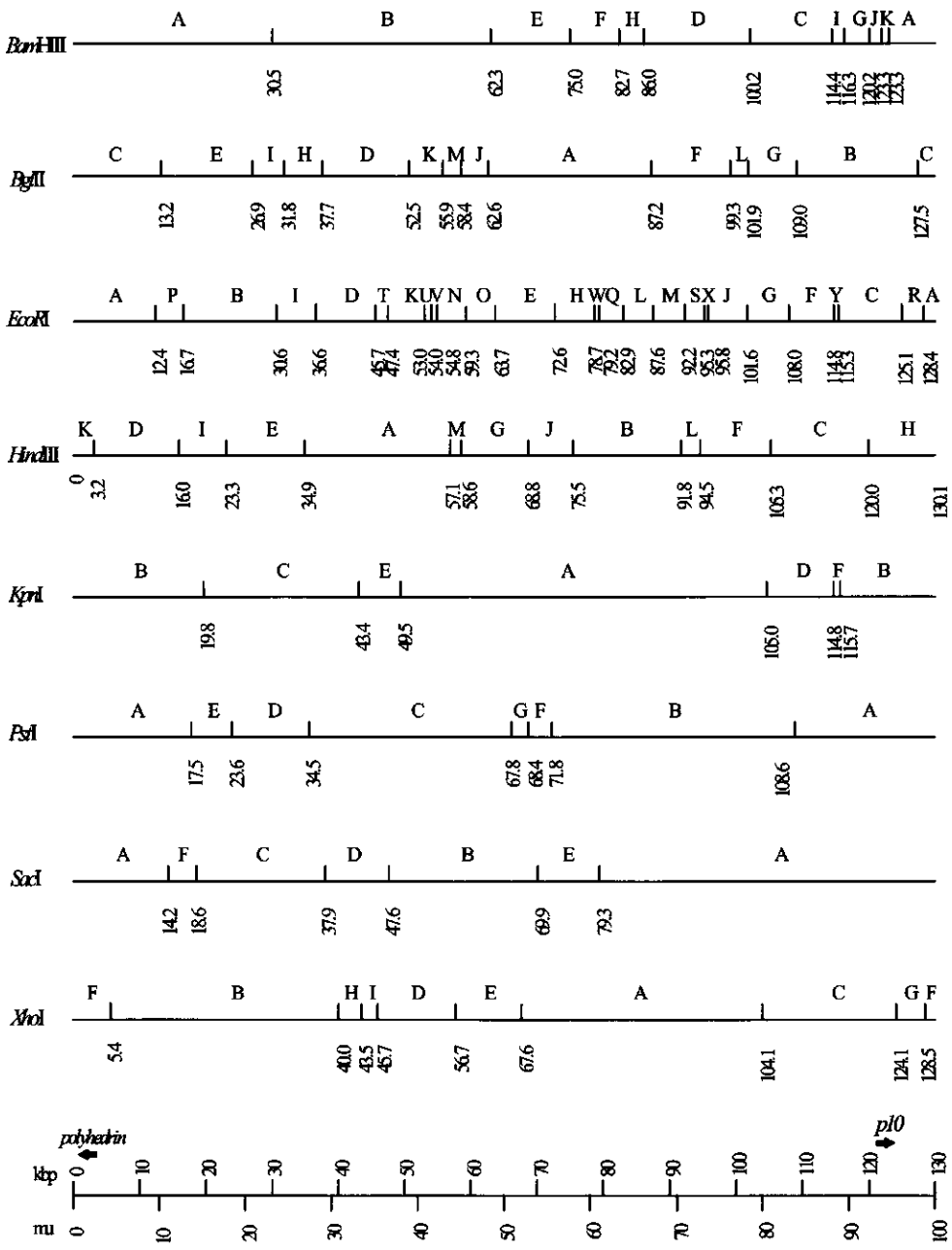


Figure 5-2. Linearized physical map of HaSNPV DNA with restriction sites for *Bam*HI, *Bgl*II, *Eco*RI, *Hin*dIII, *Kpn*I, *Pst*I, *Sac*I and *Xho*I. The restriction sites are indicated in kb from the zero point. The genome size in kb (upper) and map units (lower) is indicated on a scale at the bottom. Arrows show the location and the direction of transcription of polyhedrin and *p*10, respectively.

other enzyme digests of HaSNPV DNA were used as hybridization probes. These data, as well as single and double digestion of cloned viral DNA fragments, were used to produce a detailed physical map of HaSNPV for 90 restriction sites involving eight restriction enzymes. Since the major part of the polyhedrin gene (*ph*) is contained within the *Hin*dIII-K fragment (Chen *at al.*, 1997a), the linearized map was drawn in such a way that the zero point is the beginning of this fragment as proposed by Vlak & Smith (Vlak and Smith, 1982). The *p10* gene (Chen *et al.*, unpublished) by convention is thus positioned to the right hand end of the physical map at map position 95.

Sequence analysis of HaSNPV DNA

In order to generate sufficient information to decipher the gross gene content and arrangement in the genome of HaSNPV, approximately 45 kb of dispersed genomic sequences were determined. A total of 53 putative ORFs and gene homologs with other baculoviruses were identified by comparison with data bases. These 53 ORFs were named Ha1 to Ha53 from left to right according to their relative location on the physical map (Fig. 5-3). Table 5-3 shows the 53 putative ORFs of HaSNPV and their 48, 47 and 37 homologues in AcMNPV, SeMNPV and XcGV, respectively. HaSNPV has a few ORFs (Ha31, Ha53 and Ha34) previously thought to be unique for either AcMNPV (Ac105, Ac152) or OpMNPV (Op66), respectively.

Comparison of the overall gene content of baculoviruses provides an initial view of the genes that are retained in all genomes (Table 5-2) and thus are likely to be essential for virus replication and survival. Sequence analysis of AcMNPV, BmNPV, OpMNPV, LdMNPV and SeMNPV has so far revealed that these viruses share 60 to 70% of the ORFs. The 53 HaSNPV ORFs represented about 80% of the HaSNPV sequences determined. Ten putative ORFs have no putative homologues in other baculoviruses by pairwise comparisons. Combined with the information from AcMNPV, BmNPV, OpMNPV, BusuNPV and LdMNPV, the result of the present analysis suggests an extensive conservation of gene content in NPV genomes including SNPVs.

A small proportion of the sequenced HaSNPV regions contains putative intergenic sequences between ORFs as evidenced from the presence of baculovirus transcription motifs (data not shown). In addition, about 10 potential ORFs larger than 50 amino acids in size were identified which do not have significant homology to any sequence in Genbank. Our data indicate that, as is the case for other baculoviruses, HaSNPV provisionally encodes some unique genes, such as the recently discovered ring-finger protein (RFP) gene (Lee *et al.*, 1997). The unique ORFs occupy less than 20% of the HaSNPV sequenced region. The high conservation of the gene content and the fact that some genes are only shared by some but not all of the NPVs suggest, that collectively baculovirus genomes are a rich reservoir of common genes. This reservoir, if shared among different baculovirus genomes through recombination of genes, could provide

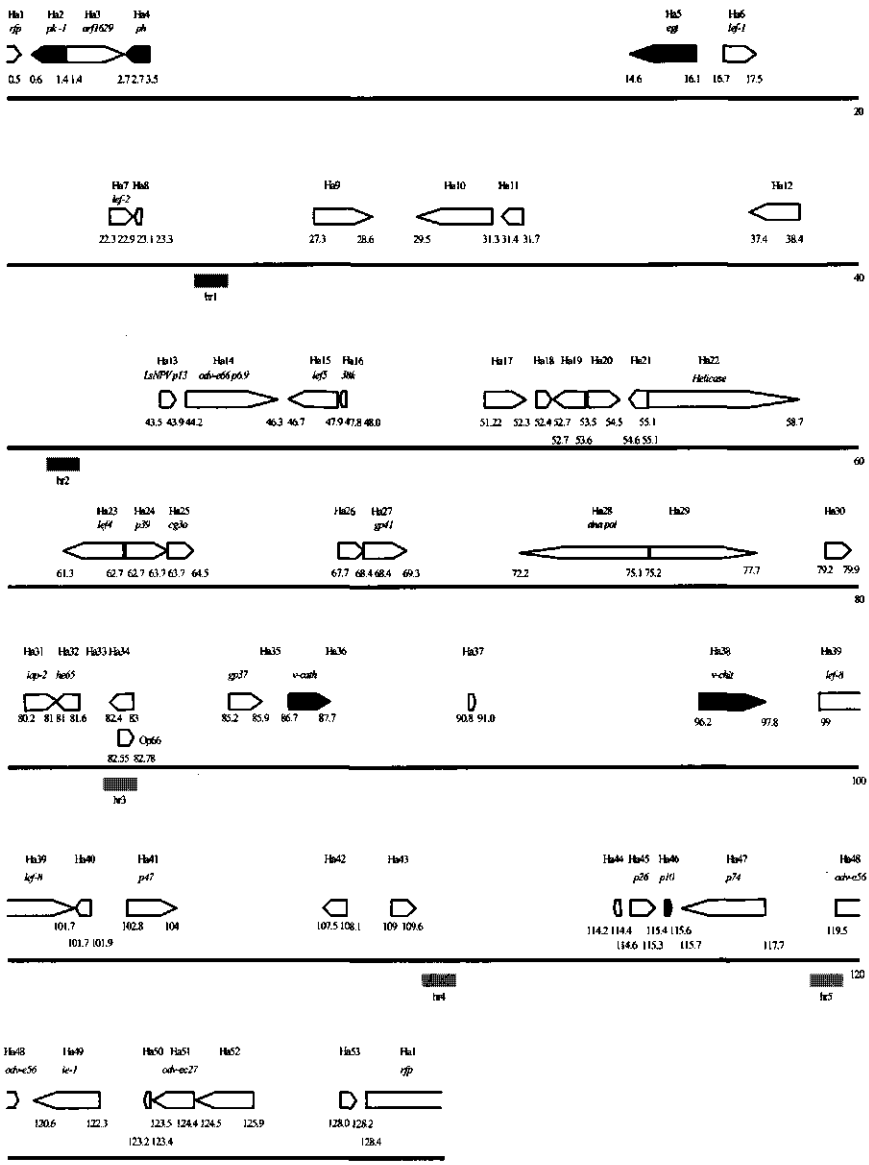


Figure 5-3. Organization of the HaSNPV genome (bars). The location of putative ORFs is shown above each bar; arrows indicate the direction of transcription; completed sequenced ORFs are shown as solid arrows. Names assigned to ORFs (Ha) are in order of appearance from the 0 point of the linearized map. The black region in each bar indicates that this section has been sequenced from both strands; the shaded region indicates a sequence from one strand; blank indicates no sequence information available. Coordinates indicating the position of the 5' and 3' end of an ORF or the sequenced region are given in kbp. The approximate locations of the *hr* regions are indicated (gray boxes).

Table 5-2. Number of homologues between eight baculoviruses after pair-wise analysis

	AcMNPV	BmNPV	OpMNPV	LdMNPV	SeMNPV	XcGV	HaSNPV	BusuNPV
AcMNPV	**154**							
BmNPV	115	**136**						
OpMNPV	126	121	**152**					
LdMNPV	94	91	95	**163**				
SeMNPV	103	99	102	104	**139**			
XcGV	84	80	76	79	72	**181**		
HaSNPV*	48	47	46	40	47	37	**53**	
BusuNPV*	47	46	48	45	45	30	26	**52**

On the basis of partial sequencing

diversity and plasticity to baculovirus species.

The high conservation of gene content is what defines a baculovirus, including the genes essential for DNA replication, gene expression and assembly of progeny viruses and occlusion at the end of the replication cycle. Many of these characteristic baculovirus genes have also been identified in the HaSNPV genome, such as those required for transient DNA replication and transcription: immediate early gene (*ie-1*), DNA helicase, DNA polymerase and late expression factor genes, *lef-1, lef-2.* But also other genes such as *lef-4, lef-5, lef-8,* as well as some structural protein genes known as *p6.9, vp39, gp41, odv-e56, odv-e66, odv-e18, odv-ec27, ph, p10, p74*, all named according to the AcMNPV designations (Ayres *et al.,* 1994) (Fig. 5-3, Table 5-3) were observed.

Apart from the above characteristic genes baculoviruses also contain so-called auxiliary genes. These genes are not essential for replication, but some provide the virus with selective advantages in nature (O'Reilly, 1997) Auxiliary genes such as those encoding EGT, inhibitor of apoptosis-like protein (IAP-2), cathepsin (CATH) and chitinase (CHI-A), protein kinase (PK-1) were also found in HaSNPV. The conservation of these auxiliary genes strongly suggests that they are important to the viral life cycle throughout the baculovirus family.

Gene arrangement of HaSNPV genome

While the gene arrangement in AcMNPV, BmNPV, and OpMNPV is basically similar, BusuNPV, belonging to Group II baculoviruses, has a distinct genome organization as compared to these three viruses (Hu *et al.*, 1998). SeMNPV and LdMNPV are also distinct in the Group II baculovirus clade in terms of genome organization. This poses the question whether the difference between MNPVs and SNPVs is also reflected by their genome organization. We therefore investigated whether the organization of the HaSNPV

genome is collinear with or different from these other viruses by subjecting the available information to GeneParityPlot analysis (Fig. 5-4). The comparison of the gene arrangement of the selected ORFs is shown in Fig. 5-4. It is clear that the gene organization in HaSNPV and SeMNPV are highly similar (Fig. 5-4b). The collinearity between HaSNPV on the one hand and AcMNPV on the other is much less, except for the central part where the collinearity among baculoviruses is very high (Fig. 5-4a) (Heldens *et al.,* 1998; IJkel *et al.*, 1999). A similar but less pronounced picture appears when HaSNPV on the one hand and BusuNPV and XcGV on the other are compared (Fig. 5-4c and d).

Attempts to find conserved gene clusters in the viral genomes may provide an initial insight into the evolution of baculovirus genomes. The methods used in this paper provide a convenient approach to identify conserved gene clusters among baculovirus genomes. Despite differences in the gene arrangement, certain genes that are clustered in the genome of AcMNPV also remained together in the HaSNPV genome. From Fig. 5-4a, nine potential clusters have been identified so far: 1. Ac 8-9-10; 2. Ac14-15; 3. Ac34-40-43; 4. Ac65-66-69; 5. Ac80-81-Ac88-89-90-Ac95-96-98-99-100-101; 6. Ac109-Ac110; 7. Ac117-119; 8. Ac136-137-138; 9. Ac142-144-145-147-148. The numbers refer to the original ORFs in AcMNPV (Ayres *et al.*, 1994) These clusters can be refined when more sequence information of HaSNPV becomes available.

In order to compare the gene arrangement between HaSNPV and BusuNPV, the 26 homologues identified so far in both genomes were chosen (Table 5-3). A comparison of the gene arrangement between HaSNPV and BusuNPV is shown in Fig. 5-4d. Although a limited number of homologues are in the equation, it is clear that the gene arrangement of HaSNPV is different from BusuNPV. Despite the overall differences in gene arrangement, there are several common gene clusters in both HaSNPV and BusuNPV: 1. Ac14-15; 2. Ac80-95-98-99; 3. Ac136-137; 4. Ac142-144-147. It is interesting to note, that besides the above common gene clusters which are the same as those in AcMNPV, these two viruses have some common characteristics. The gene cluster Ac29-136-137, which occurs in the genome of BusuNPV and HaSNPV, also exists in the genome of SeMNPV (IJkel *et al.*, 1999; Van Strien 1997). The inverted cluster encompassing Ac80-Ac101 also exists in the genome of these viruses.

Most lepidopteran NPVs fall into two large groups, Group I and Group II according to phylogenetic tree based on polyhedrin, EGT, LEF-2 and DNA polymerase (Bulach *et al.*, 1999; Chen *et al.,* 1999; Hu *et al.,* 1997; Zanotto *et al.*, 1993). AcMNPV, BmNPV, and OpMNPV are in Group I, whereas HaSNPV, BusuNPV and SeMNPV are in Group II. The gene arrangement of AcMNPV, BmNPV and OpMNPV are collinear, whereas the gene arrangements of BusuNPV and HaSNPV are different from the group I viruses as well as among each other. It seems that Group I viruses have similar gene arrangements,

Table 5-3. Identified ORFs in HaSNPV and their homologues in AcMNPV, SeMNPV, XcGV and BusuNPV

Ha ORF	Ac ORF	Se ORF	XcGV ORF	Busu ORF	Name	Abbrev	Ha dir	Ac dir	Se dir	XcGV dir	Busu dir
1		4			HzSNPV RFP				<		
2	10	3	3	52	protein Kinase	*pk-1*	<	>	>	>	<
3	9	2	2		capsid protein	*orf1629*	>	<	<	<	
4	8	1	1	1	polyhedrin	*ph*	<	>	>	>	<
5	15	27		4		*egt*	<	>	>		>
6	14	14	82	3	transcription factor	*lef-1*	>	<	<	<	<
7	6	12	35	12	transcription factor	*lef-2*	>	>	>	>	>
8	111		160	7			<	<		<	>
9	133	41	145			*alk-exo*	>	>	<	>	
10	119	36	84	8			<	>	>	>	<
11	117	47		16			<	>	<		<
12	107	53					<	>	>		
13		56	43		LsNPV p13				>	>	
14	46	57	149		ODV protein	*odv-e66*	>	>	>	<	
15	109	59	53				<	<	<	>	
16	110	60	51				<	<	<	>	
17	101	64	93				>	<	>	>	
18	100	65	94		basic protein	*p6.9*	>	<	>	>	
19	99	66	95	22	transcription factor	*lef-5*	<	>	<	<	<
20	98	67	96	23		*38k*	>	<	>	>	>
21	96	69	97				<	>	<	<	
22	95	70	98	24		*helicase*	>	<	>	>	>
23	90	74	110		transcription factor	*lef-4*	<	>	<	<	
24	89	75	111		capsid protein	*p39*	>	<	>	>	
25	88	76				*cg30*	>	<	>		
26	81	79	120				>	<	>	>	
27	80	80	121	25		*gp41*	>	<	>	>	>
28	65	93	132		DNA polymerase	*dna pol*	<	<	>	>	
29	66	92					>	>	<		
30	69	89					>	>	<		
31	71	88		30	iap-1 like	*iap-2*	<	<	>		<
32	105		67			*he65*	>	>		<	
33					LsNPV P20						
34					OpNPV ORF66						
35	64	25	107			*gp37*	>	<	>	>	
36	127	16	58	41	cathepsin	*v-cath*	>	>	<	>	>
37	59	101		36			>	<	>		<
38	126	19	103	31	chitinase	*chitA*	>	<	>	>	<
39	50	112	148	32	transcription factor	*lef-8*	>	<	>	<	<
40	43	113					<	>	<		
41	40	115	78	40		*p47*	>	<	>	>	<
42					LsNPV ec25						
43	34	124					>	<	>		
44	29	128	16	48			<	<	<	>	<
45	136	129		49		*p26*	>	>	>		>
46	137	130	5	50	fibrillin	*p10*	>	>	>	>	>
47	138	131	77	42		*p74*	<	<	<	>	<
48	148	6	15	51	ODV protein	*odv-e56*	>	<	>	<	<
49	147	132	9	46	transactivator	*ie-1*	<	>	<	<	>
50	145	134	11				<	>	<	<	
51	144	135	112	45	ODV protein	*odv-e27*	<	>	<	>	>
52	142	137	13	43			<	>	<	<	>
53	152						>	<			

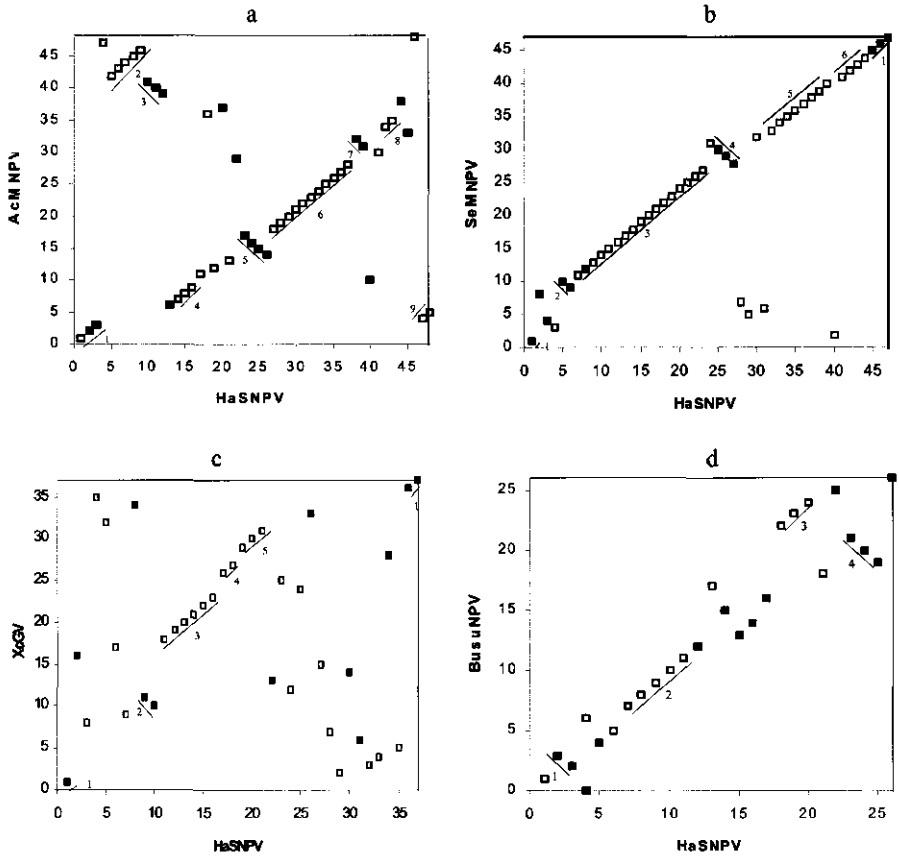


Figure 5-4. Graphic representation of the comparison of HaSNPV with the genomes of AcMNPV (a), SeMNPV (b), XcGV (c) and BusuNPV (d) obtained by GeneParityPlot analysis (see method and Table 3). Putative gene clusters of the HaSNPV genome, which are similar to that of the respective genomes, are numbered and underlined. Solid symbols indicate that the direction of the transcription of an ORF is opposite to its homologue in AcMNPV, whereas the open symbols indicate the same direction of transcription. Parallel lines mean horizontally transposed; perpendicular lines mean inverted. The polyhedrin gene is renumbered as 1; the ORFs downstream on the physical map (Fig. 2) are assigned the lowest numbers.

but the gene arrangement of Group II viruses may be much more diverse or need further subgrouping. From the present data we conclude that HaSNPV and SeMNPV cluster together within Group II or form a separate (sub)group. It has been suggested that common gene arrangements and re-arrangement in other regions on the genome may reflect the virus evolution history. Thus, gene arrangement can be used as a marker for polygenetic study of baculoviruses as proposed by Van Strien (1997).

Identification and location of HaSNPV *hr* regions

During Southern blot hybridization to construct the HaSNPV physical map, one restriction

enzyme fragment hybridized to several other fragments of the same enzyme digests (data not shown). This suggested the presence of homologous regions (*hr*s), that are common in MNPVs (Cochran and Faulkner, 1983; Possee and Rohrmann, 1997) and that serve as enhancers of transcription and, putatively, as origins of DNA replication (Guarino and Summers, 1986). One *hr*, *hr*4, was entirely sequenced and contain multiple palindromic repeats (Chen *et al.*, 2001). Five HaSNPV *hr*s were identified and approximately located on the physical map of the HaSNPV genome by hybridization with a *hr*4 repeat-specific probe (Fig. 5-3). HaSNPV *hrs* 1 through 5 were located (from left to right) on *Pst*I-D (*hr*1; 18.8-21.1 m.u.), *Xho*I-H (*hr*2; 32.0-33.1 m.u.), *Hin*dIII-B (*hr*3; 61.7-64.9 m.u.); *Kpn*I-D (*hr*4; 84.5-86.0 m.u.) and *Bam*HI-G (*hr*5; 89.5-92.5 m.u.) (Fig. 5-3). This is the first report of *hr* sequences in SNPVs to date. The nature of all these *hr*s will be confirmed by sequencing and functional analysis as enhancers of transcription and origins of DNA replication. The above results indicate that *hr*s are a common feature of baculovirus genomes.

CONCLUSION

The close relatedness between HaSNPV and SeMNPV based on gene phylogeny reconstruction (Bulach *et al.*, 1999; Chen *et al.*, 1999; Hu *et al.*, 1997; Zanotto *et al.*, 1993) is supported by the high degree of collinearity of their genomes. Since the evolutionary pressure to maintain an ORF through its functionality is different from the pressure to maintain this gene in a certain location e.g. through transcriptional constraints, one would expect incongruity between gene and genomic phylogeny. Since this is not the case, it is logical to assume that both approaches are suitable and complementary for phylogeny reconstruction. From the gene and genomic phylogeny data it can be concluded that there is no separate ancestral lineage for SNPVs and MNPVs. This suggests that changes from S to MNPVs, and vice versa, frequently occurred independently. The results thus supports the view that the baculovirus morphotype is only useful for baculovirus taxonomy and not for phylogeny.

ACKNOWLEDGEMENTS

We are grateful to Dr D. Zuidema and Prof. Peng Huiying for critical review of the manuscript. This research was supported in partly by the National Natural Science Foundation of China (NSFC 39670028, 39970031, 39980001), the Hundreds of Talents Program (STZ-3-01) and Young Scientist Group Projects of Chinese Academy of Sciences (CAS), 863 project of China, and a joint grant from CAS and the Royal Netherlands Academy of Sciences (KNAW, contract 99CDP023). The authors also gratefully acknowledge the support of the K. C. Wong Education Foundation, Hong Kong. X. Chen is a recipient of a PhD Sandwich Fellowship from Wageningen University.

Chapter 6

THE SEQUENCE OF THE *Helicoverpa armigera* SINGLE NUCLEOCAPSID NUCLEOPOLYHEDROVIRUS GENOME

SUMMARY

The nucleotide sequence of the *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus (HaSNPV) DNA genome was determined and analyzed. The circular genome encompass 131,403 bp, has a G+C content of 39.1 mol % and contains five homologous regions (*hr*) with a unique pattern of repeats. Computer-assisted analysis revealed 135 putative ORFs of 150 nucleotides or larger; 100 ORFs have homologues in *Autographa californica* multicapsid NPV (AcMNPV) and a further 15 ORFs have homologues in other baculoviruses such as *Lymantria dispar* (LdMNPV), *Spodoptera exigua* MNPV and *Xestia c-nigrum* granulovirus (XcGV). Twenty ORFs are unique to HaSNPV without homologues in GenBank. Among the six previously sequenced baculoviruses AcMNPV, *Bombyx mori* NPV, *Orgyia pseudotsugata* MNPV, SeMNPV, LdMNPV and XcGV, sixty-five ORFs are conserved and hence considered as core baculovirus genes. The mean overall amino acid identity of HaSNPV ORFs was the highest with SeMNPV and LdMNPV homologues. Other than three 'baculovirus repeat ORFs' (*bro*) and two 'inhibitor of apoptosis' (*iap*) genes no duplicated ORFs were found. A putative ORF showing similarity to poly-(ADP-ribose) glycohydrolases (*parg*) was newly identified. The HaSNPV genome lacks a homologue of the major budded virus (BV) glycoprotein gene *gp64,* of AcMNPV, BmNPV and OpMNPV. Instead, a homologue of SeMNPV ORF8, encoding the major BV envelope protein, has been identified. GeneParityPlot analysis suggests that the HaSNPV, SeMNPV and LdMNPV (Group II) have structural genomic features in common distinct from the Group I NPVs and from the granuloviruses. Cluster alignment between Group I and Group II baculoviruses suggests that they have a common ancestor.

This chapter has been published as:

Chen, X., IJkel, W. F. J., Tarchini, R., Sun, X., Sandbrink, H., Wang, H., Peters, S., Zuidema, D., Lankhorst, R. K., Vlak J. M. and Z. Hu 2001, The sequence of the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus genome. *Journal of General Virology* 82:241-257.

INTRODUCTION

The *Baculoviridae* are a family of rod-shaped viruses with circular, covalently closed, double-stranded DNA genomes ranging from 100-180 kb. The virions are occluded into large proteinaceous capsules or occlusion bodies. Two genera have been recognized: Nucleopolyhedrovirus (NPV) and Granulovirus (GV). Each genus is distinguished by a particular occlusion body morphology with single (GV) and multiple (NPV) virions occluded in granules and polyhedra, respectively. The NPVs are designated single (S) or multiple (M) depending on the potential number of nucleocapsids packaged in a virion, but this appears to have no taxonomic value (Murphy *et al.*, 1995).

Baculoviruses are frequently used as bio-insecticides of phytophagous insects, belonging mainly to the orders *Lepidoptera*, *Hymenoptera* and *Diptera* (Moscardi, 1999; Federici, 1999). The SNPV of the bollworm *Helicoverpa armigera* (HaSNPV) has been extensively used to control this insect in cotton and vegetable crops in China (Zhang, 1994). In 1999 about 100,000 hectares of cotton have been treated with a commercial virus preparation based on HaSNPV. Recently, recombinant HaSNPVs with improved insecticidal properties have been engineered (Chen *et al.*, 2000b) and field-tested (Sun *et al.*, 2001). However, the genetics of HaSNPV have only been partly described.

The nucleotide sequences of five MNPVs, notably *Autographa californica* (Ac) MNPV (Ayres *et al.*, 1994), *Bombyx mori* (Bm) NPV (Gomi *et al.*, 1999), *Orgyia pseudotsugata* (Op) MNPV (Ahrens *et al.*, 1997), *Lymantria dispar* (Ld) MNPV (Kuzio *et al.*, 1999), and *Spodoptera exigua* (Se) MNPV (IJkel *et al.*, 1999), and one granulovirus, *Xestia c-nigrum* (Xc) GV (Hayakawa *et al.*, 1999), have been determined. The size of these genomes ranges from 128,413 bp for BmNPV to 178,733 bp for XcGV. This size difference is predominantly due to the presence of gene duplications including the so-called 'baculovirus repeat ORF' or *bro* genes (Gomi *et al.*, 1999). However, no SNPV genome has been sequenced to date and it is therefore of interest to see whether the sequence of HaSNPV would reveal some unique features contributing, among others, to the SNPV phenotype and to the specificity of this virus for heliothine insects.

A physical map of HaSNPV has been constructed previously and the size was estimated to be about 130 kb (Chen *et al.*, 2000a). Analysis of approximately 45 kb of random sequence from the HaSNPV genome resulted in the identification of 53 open reading frames with homologies to ORFs of other baculoviruses. Partial alignment of the HaSNPV genome with other baculovirus genomes using GeneParityPlot (Hu *et al.*, 1998) revealed a close relationship of HaSNPV and SeMNPV in terms of genomic organization (Chen *et al.*, 2000a). A few genes, notably polyhedrin (*ph*) (Chen *et al.*, 1997b), the ecdysteroid UDP glucosyltransferase (*egt*) (Chen *et al.*, 1997a), DNA polymerase (Bulach *et al.*, 1999) and 'late expression factor 2' (*lef-2*) (Chen *et al.*, 1999), have been characterized in

some detail. Phylogenetic analysis of these genes also revealed a close ancestral relationship between HaSNPV, SeMNPV and LdMNPV at the gene level.

In this paper we describe the complete nucleotide sequence and organization of the *Helicoverpa armigera* SNPV genome. This baculovirus is characterized by the absence of extensive gene duplications and by the presence of a limited number of homologous repeat (*hr*) regions, whose structure is distinctly different from the *hr* sequences of other baculoviruses. Finally, a genomic comparison is made with the complete sequences of AcMNPV, SeMNPV, LdMNPV and XcGV using GeneParityPlot (Hu *et al.*, 1998).

MATERIALS AND METHODS

Insect and virus

The bollworm *H. armigera* was cultured as a laboratory colony and reared on artificial diet as described by Zhang *et al.* (1981). The wild type virus was originally isolated from diseased *H. armigera* larvae in Hubei province, People's Republic of China in 1981. By *in vivo* cloning eight HaSNPV genotypes were isolated (G1-G8) (Sun *et al.*, 1998), of which the G4 strain was selected for sequencing. Polyhedra of the G4 strain were propagated in fourth instar *H. armigera* larvae.

HaSNPV DNA isolation, cloning, and sequence determination

The HaSNPV G4 strain (Sun *et al.*, 1998) was sequenced to a 6-fold genomic coverage using a shotgun approach. The viral DNA was cesium chloride-purified and sheared by nebulization into fragments with an average size of 1,200 bp. Blunt repair of the ends was performed with *Pfu* DNA polymerase (Stratagene) according to the manufacturer's directions. DNA fragments were size-fractionated by gel electrophoresis and cloned into the *Eco*RV site of pBluescriptSK (Stratagene). After transformation into XL2 blue competent cells (Stratagene) a thousand recombinant colonies were picked randomly. DNA templates for sequencing were isolated using QIAprep Turbo kits (Qiagen) on a QIAGEN BioRobot 9600. Sequencing was performed using the ABI PRISM Big Dye Terminator Cycle Sequencing Ready reaction kit with FS AmpliTaq DNA polymerase (Perkin Elmer) and analyzed on an ABI 3700 DNA Analyzer.

Shotgun sequences were base-called by the PRED basecaller, and assembled with the PHRAP assembler (Ewing *et al.*, 1998; Ewing & Green, 1998). Using the PREGAP4 interface PHRAP-assembled data were stored in the GAP4 assembly database (Bonfield *et al.*, 1995). The GAP4 interface and its features were then used for editing and sequence finishing. Consensus calculations with a quality cutoff value of 40 were performed from within GAP4 using a probabilistic consensus algorithm based on expected error rates output by PHRED. Sequencing PCR products bridging the ends of existing contigs filled

remaining gaps in the sequence.

DNA sequence analysis

Genomic DNA composition, structure, repeats, and restriction enzyme pattern were analyzed with the University of Wisconsin Genetics Computer Group programs (Devereux *et al.*, 1984) and DNASTAR™ (Lasergene). Open reading frames (ORFs) encoding more than 50 amino acids (150 bp) were considered to be protein encoding and hence designated putative genes. One hundred and fifteen ORFs (out of 135) were checked for maximal alignment with known baculovirus gene homologues extracted from GenBank; ORFs with overlap of *hrs* were excluded from the alignment analysis. The overlap between any two ORFs with known baculovirus homologues was set to a maximum of 25 amino acids; otherwise the largest ORF was selected.

DNA and protein comparisons with entries in the sequence databases were performed with FASTA and BLAST programs (Pearson, 1990; Altschul *et al.*, 1990). Multiple sequence alignments were performed with the GCG PileUp and Gap computer programs (Genetics Computer Group Inc., Madison, Wisconsin, version 10.0) with gap creation and extension penalty set to 9 and 2, respectively (Devereux *et al.*, 1984). Percent identity indicates the percent identical residues between two complete sequences. The GENESCAN program was used for gene predictions. The DOTTER program was used to identify and classify repeat families and MITEs (Miniature Inverted Repeat Transposable Element). GeneParityPlot analysis was performed on the HaSNPV genome versus the genomes of AcMNPV, SeMNPV, LdMNPV and XcGV as described previously (Hu *et al.*, 1998).

Nucleotide sequence accession number.

The HaSNPV genome sequence has been deposited in GenBank under accession no. AF271059.

RESULTS AND DISCUSSION

Nucleotide sequence analysis of the HaSNPV genome

The HaSNPV genome was assembled into a contiguous sequence of 131,403 bp (Table 6-1). This size is in good agreement with a previous estimate of 130.1 kb for HaSNPV DNA based on restriction enzyme analysis and physical mapping (Chen *et al.*, 2000a). AcMNPV, BmNPV, OpMNPV and SeMNPV have similar size genomes, which are much smaller than the genomes of LdMNPV and XcGV with 161 kb and 178 kb respectively (Table 6-1). With a G+C content of 39.1 % HaSNPV has the lowest G+C content among baculoviruses to date, which is close to that of AcMNPV (41%) (Ayres *et al.*, 1994),

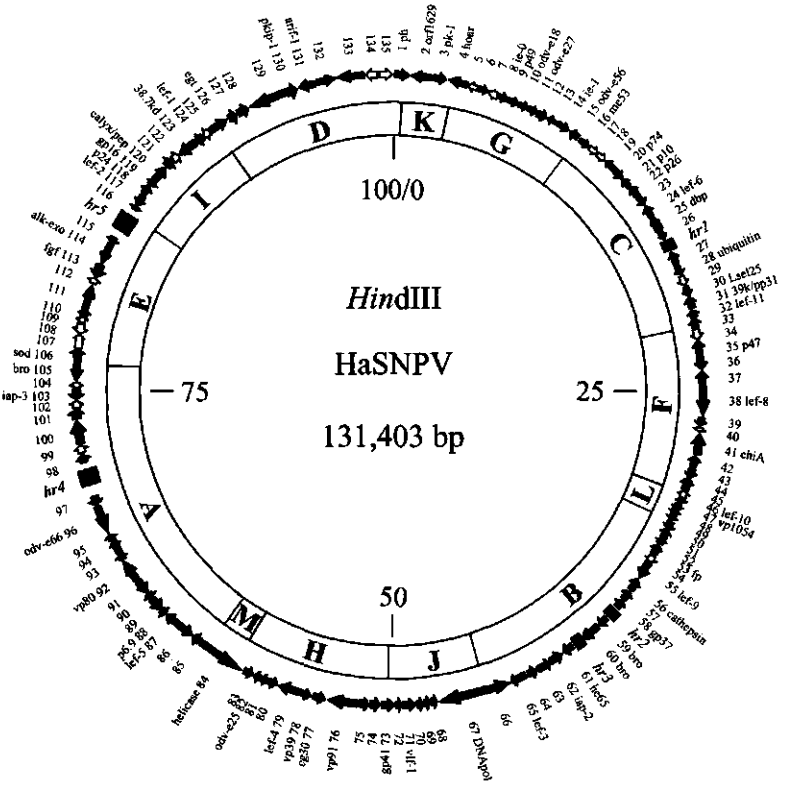


Figure 6-1 Circular map of the HaSNPV DNA genome. Genomic organization of the circular HaSNPV genome. The sites for restriction enzyme *Hin*dIII are presented; the fragments are indicated A to M according to size from the largest to the smallest restriction fragment (Chen *et al.*, 2000a). The position of the 135 identified ORFs is indicated with arrows that also represent the direction of transcription. Shaded arrows indicate that the ORF has a homologue in other baculoviruses in the protein sequence databases. Open arrows represent ORFs unique to HaSNPV. The corresponding number along the ORF represents the HaSNPV ORF number. The position of the *hr* sequences is indicated by black boxes. The scale on the inner circle is in map units.

BmNPV (Gomi *et al.*, 1999) and XcGV (Hayakawa *et al.*, 1999). The G+C content of OpMNPV (Ahrens *et al.*, 1997) and LdMNPV (Kuzio *et al.*, 1999) are much higher with 55 mol % and 58 mol %, respectively. According to a recently adopted convention (IJkel *et al.*, 1999; Hayakawa *et al.*, 1999) the adenine residue at the translational initiation codon of the polyhedrin gene was designated as the zero point of the physical map of HaSNPV DNA (Fig. 6-1). Taking polyhedrin as the first gene this determines the orientation of the physical map. This map is now reversed as compared to the original map presented by Chen *et al.* (2000a) and positions the p10 gene at map unit 10.

By computer-assisted analysis 326 ORFs defined as methionine-initiated ORFs larger than 50 amino acids were found. From these, one hundred and thirty five ORFs with less than 25 amino acids or no overlap with other ORFs have been identified on the HaSNPV genome (Fig. 6-1; Tables 6-1 and 6-2) and were further analyzed. This number of 135 ORFs is roughly proportional to the size of the HaSNPV genome as compared to the other six completely sequenced baculovirus genomes AcMNPV, BmNPV, OpMNPV, SeMNPV, LdMNPV and XcGV. The HaSNPV ORFs are in general tightly packed with minimal intergenic distances; their orientation is almost evenly distributed along the genome (52% clockwise, 48% counterclockwise; Fig. 6-1). The location, orientation and size of the predicted ORFs are shown in detail in Table 6-2. The 135 predicted ORFs account for 87% of the genome versus 8% for intergenic sequences and 6% for *hr* regions. The HaSNPV ORFs have an average length of 844 nucleotides with Ha84 (*helicase*) being the largest (3,758 nt) and Ha40, without a homologue in other baculoviruses, being the smallest (150 nt). One hundred and fifteen (86%) of the 135 HaSNPV ORFs have an

Table 6-1. Characteristics of different baculovirus genomes. The genome characteristics of the different baculoviruses are derived from the following reference AcMNPV (Ayres *et al.*, 1994), BmNPV (Gomi *et al.*, 1999), OpMNPV (Ahrens *et al.*, 1997), LdMNPV (Kuzio *et al.*, 1999), SeMNPV (IJkel *et al.*, 1999), XcGV (Hayakawa *et al.*, 1999)

	HaSNPV	AcMNPV	BmNPV	OpMNPV	LdMNPV	SeMNPV	XcGV
size (kb)	131.4	133.9	128.4	132.0	161.0	135.6	178.7
G+C (%)	39	41	40	55	58	44	41
total ORFs	135	154	136	152	163	139	181
unique ORFs	20	11	1	16	29	17	82
hrs	5	8	7	5	13	6	8
early	33	65	12	61	12	34	13
late	60	72	78	64	79	72	84
early + late	9	29	7	26	6	14	2
promoter not identity	49	47	35	58	78	53	84

assigned function or have homologues with other putative baculovirus genes (Table 6-2). Twenty ORFs are so far unique to HaSNPV. These ORFs accounted for 6% (7.3 kb) of the genome in total.

The HaSNPV nucleotide sequence was determined from an isolate cloned *in vivo* (Sun *et al.*, 1998). Based on restriction enzyme analysis and Southern hybridization no fragments in a less than molar ratio were observed in this isolate. However, sequence analysis showed that at approximately 100 nucleotide locations (0.07% of the genome) along the genome a polymorphism was observed in the nucleotide usage. None of these affected ORFs. This polymorphism may be partly the result of the sequencing (error 10^{-5}), but also the consequence of the intrinsic genetic variation, that exists in natural HaSNPV isolates (Gettig & McCarthy, 1982; Figueiredo *et al.*, 1999) or *in vivo* cloned isolates of HaSNPV (Sun *et al.*, 1998), GV (Smith & Crook, 1988) and MNPVs (Muñoz *et al.*, 1998). Despite the *in vivo* cloning and the apparent lack of genetic heterogeneity as evidenced from restriction enzyme analysis (Sun *et al.*, 1998) microvariation may thus exist. This suggests that the quasispecies concept for RNA viruses, i.e. a virus species is defined not as a single nucleotide sequence but as a mixture of genotypes (Domingo *et al.*, 1995), may also apply for DNA viruses including baculoviruses.

Homologous regions (*hrs*)

Regions with homologous repeats were first found in AcMNPV (Cochran & Faulkner, 1983) and appear to be present in all baculoviruses. They occur at multiple locations along the genome and may serve as origins of DNA replication (Kool *et al.*, 1995) and as enhancers of transcription (Guarino & Summers, 1986; Guarino *et al.*, 1986). *Hrs* are characterized by the presence of multiple, often imperfect tandemly repeated palindromic sequences (AcMNPV). Five *hr* regions were identified previously on the genome of HaSNPV by direct sequencing and Southern blot hybridization (Chen *et al.*, 2000a). No further *hrs* were detected in the complete sequence (Fig. 6-1; Table 6-1)

The five *hrs* were found dispersed along the HaSNPV genome around map positions 17.5 (*hr1*), 37.7 (*hr2*), 40.2 (*hr3*), 70.8 (*hr4*) and 83.6 (*hr5*) and are located in AT-rich intergenic regions. Their sizes vary from 750 (*hr3*) to 2800 nucleotides (*hr5*). It is interesting to note that *hr2* and *hr3* are separated by two *bro*-related genes (Fig. 6-1). This configuration might have been the result of an insertion of two *bro* genes into what originally may have been a single *hr*. Assuming *hr2* and *hr3* had been a single *hr*, the *hrs* of HaSNPV are remarkably similar is size (2100 - 2800 nt).

Using a dot matrix analysis the HaSNPV sequence was compared to itself and its complementary strand. Two types of repeats were identified, type A and type B, with imperfect 40 bp-long and 107 bp-long repeats, respectively, or truncated versions

Table 6-2. Listing of potentially expressed ORFs in HaSNPV

HaSNPV ORF								Homologous ORFs						Identity to homologues (%)							cluster
ORF	Name	left	Right	Dir	Aa	Mr	prom	Ac	Bm	Op	Ld	Se	Xc	Ac	Bm	Op	Ld	Se	Xc	Hz*	
1	*ph*	1	738	>	245	26779	L	8	1	3	1	1	1	85	81	83	80	86	53	99	a
2	*orf1629*	735	1976	<	413	45905		9	2	2	2	2	2	30	27	24	28	27	29	99	a
3	*pk-1*	1991	2794	>	267	31543		10	3	1	3	3	3	41	41	40	43	55	36	100	a
4	*hoar* *	2917	5187	<	756	85428	E					4						27		97	
5		5383	5562	>	59	7219															
6	*Hzorf480*	5733	6590	>	285	34459	E													100	
7		6794	6961	<	55	6436															
8	*ie-0*	6949	7806	>	285	33186	L L	141	117	138	21	138		31	31	32	33	31			b
9	*p49*	7823	9229	>	468	55256	L L L	142	118	139	20	137	13	52	52	51	56	57	35	99	b
10	*odv-e18*	9240	9485	>	81	8822	L	143	119	140	19	136	12	62	56	44	58	68	50	90	b
11	*odv-ec27*	9500	10354	>	284	33288	L	144	120	141	18	135	112	52	52	50	56	59	32	100	b
12		10399	10677	>	92	10780	L	145	121	142	17	134	11	50	48	49	58	58	23		b
13		10704	11315	<	203	22922		146	122	144	16	133	10	32	31	32	31	34	26	92	b
14	*ie-1*	11357	13324	>	655	75972		147	123	145	15	132	9	30	30	30	34	34	22	98	b
15	*odv-e56*	13378	14442	<	354	38850	L L	148	124	146	14	6	15	49	49	50	50	51	44	100	b
16	*me53*	14603	15457	>	284	33603	E	139	116	137	23	7	180	23	24	25	33	33	27		
17		15504	15683	>	59	7340															
18		15686	15853	>	55	6377															
19		15906	16187	<	93	11110	E				26						40				
20	*p74*	16208	18274	>	688	78434		138	115	134	27	131	77	53	54	54	59	57	41		
21	*p10*	18328	18591	<	87	9331	L	137	114	133	41	130	5	26	32	23	47	48	59		c
22	*p26*	18674	19477	<	267	30510	L	136	113	132	40	129		38	37	34	33	23			c
23		19591	19794	>	67	8258	E	29	20	39	39	128	16	32	31	33	45	44	39		
24	*lef-6*	19870	20433	<	187	22188		28	19	40	38	127		30	30	30	37	38			d
25	*dbp*	20447	21421	<	324	37560		25	16	43	47	126	89	36	37	31	24	49	25		d
26		21638	22039	>	133	15025		26	17	42	36	125		35	32	35	39	29			
[illegible]																					
27		24316	25083	<	255	29529	E	34	25	26	42	124		31	33	35	46	54			e
28	*ubiquitin*	24923	25174	>	83	9244	L L	35	26	25	43	123	52	73	73	74	76	74	79		e
29		25238	25744	>	168	20406	E														
30	*Lsel25* **	25764	26336	>	190	22541	L													31**	
31	*39K / 9931*	26395	27330	<	311	35195		36	27	24	44	120	55	36	36	33	40	36	24		e
32	*lef-11*	27296	27679	<	127	14583		37	28	23	45	119	56	35	35	31	41	48	35		e
33		27648	28364	<	238	28411		38	29	22	46	118	79	53	53	57	57	61	47		e
34		28595	29674	>	359	41190	E														
35	*p47*	29985	30986	<	333	38963		40	31	45	48	115	78	53	53	47	58	56	42		
36		31059	31730	>	223	25768	E	41	32	46				26	25	26					
37		31816	32058	>	80	9543	L	43	34	48		113		25	25	25		31			
38	*lef-8*	32055	34760	<	901	1E+05		50	39	54	51	112	148	64	65	60	67	70	54		
39		34813	35397	>	194	22508	L	51	40	55		111		26	24	25		26			
40		35538	35690	>	50	6299															
41	*chitinase*	35698	37410	<	570	65481		126	103	124	70	19	103	67	68	68	66	63	60	95	
42		37489	38031	<	180	21260	E	52	41		53	109		29	31		38	26			f
43		38148	38558	>	136	16419		53	42	56	54	108	171	42	44	45	49	56	28		f
44		38565	39701	<	378	42771	L				55	107					26	30			f
45		39709	39936	<	75	9090	L														
46	*lef-10*	39896	40111	>	71	7684		53a	42a	57	56	106	171	43	42	31	43	54	36		f
47	*vp1054*	39984	41039	>	351	41700		54	43	58	57	105	175	44	44	40	50	55	40		f
48		41159	41365	>	68	7962		55	44	59	58	104		40	31	31	40	53			f
49		41366	41560	>	64	7406	L	56	45	60		103		26	26	28		39			f
50		41846	42361	>	171	20671	L	57	46	61	60	102		42	42	41	44	44			f

HaSNPV ORF								homologous ORFs						Identity of homologues(%)							cluster
ORF	Name	left	right	dir	Aa	Mr	pro	Ac	Bm	Op	Ld	Se	Xc	Ac	Bm	Op	Ld	Se	Xc	Hz	
51		42412	42894	<	160	19034		59		62	61	101		28		38	39	47			f
52		42906	43172	<	88	10219	L	60	48	63	62	100	102	43	44	31	45	57	43		f
53	*fp*	43385	44038	<	217	25368	L	61	49	64	63	98	140	62	62	56	52	68	37		f
54		44210	44395	>	61	7302															
55	*Lef-9*	44507	46066	>	519	59545		62	50	65	64	97	139	65	66	53	70	72	57		f
56	*v-cath*	46150	47247	<	365	42021	L	127	104	125	78	16	58	47	48	48	47	46	44		
57		47288	47875	<	195	21292	E L						83						33		
58	*Gp37*	47946	48785	<	279	32099	E L	64	52	69	68	25	107	56	56	56	58	60	45		
	hr2																				
59	*Bro-a*	49936	50670	>	244	28269					150						53				
60	*Bro-b*	50794	52377	>	527	59734					146		159				60		58***		
	hr3																				
61	*He65*	53133	53843	>	236	27478	E	105	89				67	29	28				33		
62	*Iap-2*	53920	54672	<	250	29254	L	71	58	74	79	88		34	35	35	41	42			g
63		54720	55544	<	274	31562		69	57			89		42	43			48			g
64		55513	55914	<	133	15561	E	68	56	73	80	90	135	42	43	35	52	56	30		g
65	*lef-3*	55934	57073	>	379	44018		67	55	72	81	91	134	27	29	29	29	35	17		g
66		57181	59538	<	785	88881	L	66	54	71	82	92		28	25	24	25	27			g
67	*DNA pol*	59569	62631	>	1020	1E+05		65	53	70	83	93	132	47	47	44	55	61	38		g
68		62708	63166	<	152	17612	L L	74	60	77				26	26	17					
69	*Hz384*	63232	63615	<	127	14880	L	75	61	78	84	94	126	24	24	25	40	38	26	100	h
70		63621	63878	<	85	9958	L	76	62	79	85	95	125	43	42	39	74	64	37		h
71	*vlf-1*	63919	65157	<	412	47878	L	77	63	80	86	82	123	70	69	67	71	64	35	99	i
72		65170	65502	<	110	12730	L E	78	64	81	87	81	122	44	41	41	44	50	33		I
73	*gp41*	65571	66539	<	322	36579	L	80	66	83	88	80	121	58	57	53	64	56	37		i
74		66469	67194	<	241	27681	E	81	67	84	89	79	120	54	53	50	55	58	52		i
75		67067	67744	<	225	24912	E	82	68	85	90	78	119	31	31	23	44	49	31		i
76	*vp91*	67674	70124	>	816	93527	L	83	69	86	91	77	118	43	43	42	42	48	33		i
77	*cg30*	70252	71103	<	283	32325	L	88	71	89		76		31	29	25		24			i
78	*vp39*	71192	72073	<	293	33403		89	72	90	92	75	111	45	46	48	52	54	35		i
79	*lef-4*	72072	73457	>	461	53977		90	73	91	93	74	110	46	46	40	46	53	37		i
80		73510	74274	<	254	30849		92	75	93	94	73	101	55	56	56	51	60	45		i
81		74276	74764	>	162	19065	L	93	76	94	95	72	100	54	54	51	61	62	36		i
82	*odv-e25*	74810	75502	>	230	25933		94	77	95	96	71	99	44	43	40	74	69	51		i
83		75534	76031	<	165	18793	E L					68						26			
84	*helicase*	76050	79811	<	1253	1E+05		95	78	96	97	70	98	44	44	40	49	50	30		I
85		79768	80289	>	173	19805	E	96	79	97	98	69	97	48	48	45	61	61	36		I
86		80348	81313	<	321	37930		98	82	99	99	67	96	46	45	45	50	55	41		k
87	*lef-5*	81209	82156	>	315	37040		99	83	100	100	66	95	52	52	50	50	57	43		k
88	*p6.9*	82150	82479	<	109	11522		100	84	101	101	65	94	44	49	53	69	59	57		k
89		82544	83653	<	369	42553	L	101	85	102	102	64	93	43	42	39	43	53	26		k
90		83699	84067	<	122	13830	L	102	86	103	103	63	92	26	26	32	26	33	26		k
91		84067	85200	<	377	44040	E L	103	87	104	104	62	91	51	51	45	52	60	41		k
92	*vp80*	85295	87112	>	605	69719	E L	104	88	105	105	61		23	24	22	26	29			k
93		87109	87285	>	58	6943		110			106	60	51	29			46	48	43		
94		87300	88385	>	361	41508		109	92	109	107	59	53	58	58	54	60	58	35		l
95		88431	88715	>	94	10974		108	91	108	108	58		41	42	34	45	51			l
96	*odv-e66*	88782	90800	<	672	76093	L	46	37	50	131	57/114	149	43	42	43	54	45/34	60		
97	*p13* ****	90821	91651	<	276	32453	L					56	43					59	48		
	hr4																				
98		93957	94556	>	199	22409	L	115	95	115	143	50	32	39	39	40	47	45	40		18
99		94560	94916	>	118	14449	E														
100	*parg*	95012	96544	>	510	58136					141	52					24	24			
101		96623	97384	>	253	29046	L	106/107	90	107	140	53	50	47	47	47	50	56	31		
102		97399	97731	>	110	12790															

HaSNPV ORF								homologous ORFs						Identity to homologues (%)							Cluster
ORF	Name	left	right	Dir	Aa	Mr	pro	Ac	Bm	Op	Ld	Se	Xc	Ac	Bm	Op	Ld	Se	Xc	Hz*	
103	*iap-3*	97789	98595	<	268	31522	E L			35	139	110				41	29	38			
104		98592	98747	<	51	5931															
105	*bro-c*	98858	1E+05	<	501	58269	L				71		60				51		66		
106	*sod*	100531	1E+05	>	159	16853		31	23	29	145	48	68	72	72	71	68	68	57		
107		101017	1E+05	>	457	51209															
108		102443	1E+05	<	192	22772															
109		103190	1E+05	>	118	13648															
110		103557	1E+05	>	88	10079		117	96	117		47		33	30	24		48			
111		103891	1E+05	>	528	60289		119	97	119	155	36	84	49	49	47	47	46	36		
112		105474	1E+05	>	78	9090	L														
113	*fgf*	105733	1E+05	<	301	34358	E	32	24	27	156	38	144	29	31	29	29	27	23		
114	*alk-exo*	106765	1E+05	<	428	49416		133	110	131	157	41	145	45	45	44	42	44	42		
115		108071	1E+05	<	129	15332	L	19	11	18	159	42		28	27	27	29	32			
[illegible]																					
116		111267	1E+05	>	71	8204	E	111	93	112	76		160	36	34	35	32		59		
117	*lef-2*	111600	1E+05	<	241	27811	E	6	135	6	137	12	35	43	43	42	42	46	29		
118	*p24*	112687	1E+05	>	248	28373	L	129	106	127		10	80	37	40	37		51	23		
119	*gp16*	113495	1E+05	>	94	10669	L L	130	107	128		9		25	25	22		32			
120	*caly*	113831	1E+05	>	340	39058	L	131	108	129	136	46		36	38	29	38	47			
121		114932	1E+05	>	154	18472	E	63	51		117			29	29		23				
122		115527	1E+05	>	196	23477	E L														
123	*38.7kd*	116174	1E+05	<	385	44474		13	5	12	122	13		26	26	25	29	38			m
124	*lef-1*	117333	1E+05	<	245	29059		14	6	13	123	14	82	39	40	44	48	51	45		m
125		118045	1E+05	<	144	16114	E L														
126	*Egt*	118624	1E+05	>	515	58870	L	15	7	14	125	27		47	47	45	50	55		99	n
127	*orf****	120371	1E+05	>	192	22595														24	
128		120900	1E+05	>	266	30352		17	9	16	128	29		26	29	30	31	33			n
129		121781	1E+05	<	947	1E+05	L L				129	30					30	29			n
130	*pkip-1*	124989	1E+05	>	169	20282		24	15	44	110	32		23	26	27	35	38			
131	*arif-1*	125565	1E+05	<	265	30355	E	21	12	19	118	34		29	24	24	25	30			
132		126619	1E+05	>	383	44534		22	13	20	119	35	45	60	60	57	66	66	50		
133		127811	1E+05	<	677	78241	L E	23	14	21	130	8	27	24	23	25	43	40	29		
134		129986	1E+05	<	182	21930	E														
135		130713	1E+05	>	194	23310	E														

* Taken from HzSNPV of GenBank

** LsNPV ORF name taken from Wang *et al.* (1995). Percentage amino acid identity is shown to Lsel25

*** Identity to C-terminal 344 amino acids

**** *S. litoralis* ORF homologue

(Fig. 6-2). The type A and B repeats are found in each of the *hrs*. There is no sequence homology with other known baculovirus *hrs*. The type B repeats contain short internal stretches of palindromic and direct repeats. Not only is the sequence of the HaSNPV *hr* regions different from those of other baculovirus *hrs*, but also the structure is rather unique. The function of the type A and type B repeats remains to be determined.

Comparison of the gene content of HaSNPV and other baculoviruses

The sequence of the HaSNPV genome was compared with those of AcMNPV (Ayres *et al.*, 1994), BmNPV (Gomi *et al.*, 1999), OpMNPV (Ahrens *et al.*, 1997), SeMNPV (IJkel *et al.*, 1999), LdMNPV (Kuzio *et al.*, 1999) and XcGV (Hayakawa *et al.*, 1999) for the

presence or absence of putative ORFs (Table 6-2). These seven baculovirus genomes have a cumulative total of 354 different ORFs, of which 183 are unique to individual baculovirus genomes. Sixty-five ORFs are conserved among the seven baculoviruses including HaSNPV. Among all NPVs 84 ORFs are conserved (data not shown). This suggests that about 70 ORFs are the minimal requirement for basic baculovirus features, such as virus structure, transcription, DNA replication, auxiliary functions on the cellular or organismal level and occlusion body morphogenesis (Table 6-2). Putative functions have been assigned to approximately 61% of these common baculovirus genes. Twenty ORFs larger than 50 amino acids were unique to HaSNPV. Nine of these, 50 to 100 amino acids long, have no consensus baculovirus promoter (Table 6-2). Most likely, these small ORFs in HaSNPV are not functional, but this has to be tested experimentally.

Of the 135 HaSNPV ORFs identified, 100 have homologues in AcMNPV and a further 15 have homologues in other baculoviruses (Tables 6-1 and 6-2). HaSNPV shares the largest number of homologues (103) with SeMNPV underscoring the close relationship between these two viruses as evidenced from gene phylogeny analyses involving *polyhedrin, egt, lef-2* (Chen *et al.*, 1997a, b; Chen *et al.*, 1999) and *DNA polymerase* (Bulach *et al.*, 1999). *Polyhedrin* is the most conserved ORF of the six NPVs, with a mean amino acid identity of 83% with other NPV *polyhedrins*; the identity to GV *granulin* is much less (51% for XcGV). *Ubiquitin* (*ubi*), which is involved in the targeting of proteins for degradation, is the next most conserved gene among the seven sequenced baculoviruses with 75% amino acid identity followed by *superoxide dismutase* (*sod*) with 70%. The mean ORF amino acid identity between HaSNPV and the group II baculoviruses SeMNPV and LdMNPV is similar (46%) and higher than to group I baculoviruses (41%). This is in support of the distinct phylogenetic relationship between group I and group II NPVs (Zanotto *et al.*, 1993; Bulach *et al.*, 1999).

Table 6-3. Number of ORFs with homologues in baculoviruses and percentage amino acid identity. The number of ORFs with homologues in baculoviruses are shown above the diagonal and the percentage amino acid identity is shown in italics below the diagonal

	AcMNPV	BmNPV	OpMNPV	LdMNPV	SeMNPV	XcGV	HaSNPV
HaSNPV	**100**	**98**	**94**	**94**	**103**	**69**	-
XcGV	84	80	76	93	72	-	*40*
SeMNPV	103	99	102	104	-	*nd*	*47*
LdMNPV	94	91	95	-	*45*	*nd*	*46*
OpMNPV	126	121	-	*nd*	*40*	*34*	*41*
BmNPV	115	-	*55*	*nd*	*41*	*nd*	*41*
AcMNPV	-	*93*	*56*	*41*	*41*	*33*	*41*

ND, Not determined

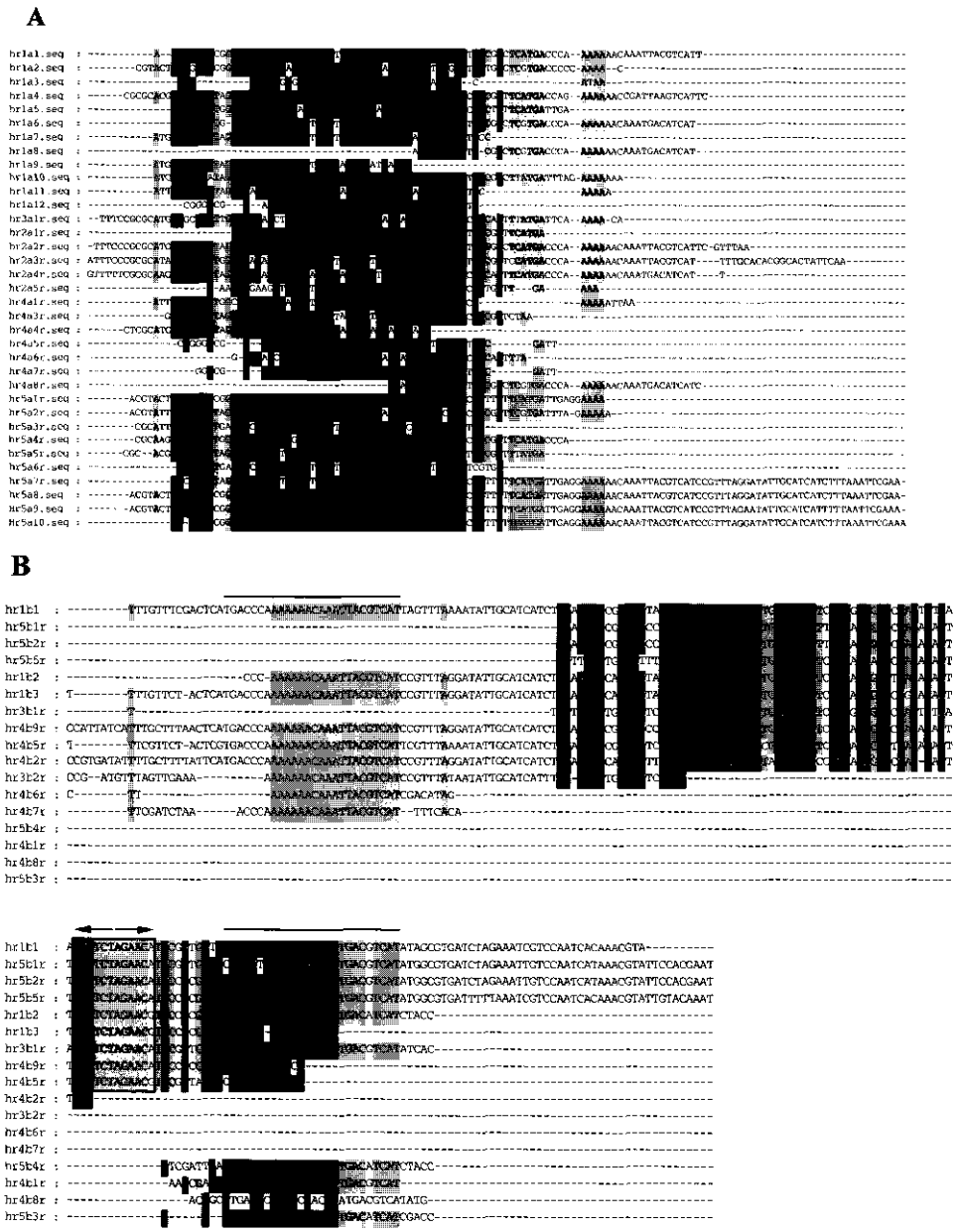


Figure 6-2. Alignment of HaSNPV repeated sequences. The nucleotide sequences of the type A repeats (panel A) and type B repeats (panel B) are aligned to obtain maximum homology. The repeats are denoted according to their presence in a homologous region (*hr*1-*hr*5), their type (a or b), their order number in the hr (1-n) and whether they occur in the reverse orientation (r) or not. Shading is used to indicate the relative occurrence of similar nucleotides in the repeats: black indicates > 59%, grey 53 % and white < 47% representation. Short palindromes are indicated (by arrows) as well as are direct repeats (lines above) and *Xba*I sites (boxed).

Structural virion genes

The HaSNPV genome contains genes known to encode the common virion structural proteins of NPVs (Table 6-2). In contrast to SeMNPV, where *odv-e66* is duplicated (IJkel *et al.*, 1999), HaSNPV does not contain duplicate genes for virion structural proteins. However, HaSNPV apparently lacks a homologue of the budded virus envelope surface glycoprotein gene, *gp64* (Ac128). The product of this gene, GP64, is acquired by virions during budding through the plasma membrane and is involved in the association with cell receptors upon invasion and fusion in endosomes (Oomens & Blissard, 1999). However, an ORF has been identified in HaSNPV (Ha133) with an average amino acid identity with Ld130 (43%) and Se8 (40%). The latter viruses also lack a *gp64* homologue and it has been suggested that Ld130 and Se8 are the functional homologues of AcMNPV *gp64* (Kuzio *et al.*, 1999; IJkel *et al.*, 1999). Recently, direct evidence was obtained that the products of Ld130 and Se8 are the major constituents of the BV envelope and responsible for the fusogenic activity of SeMNPV (Pearson *et al.*, 2000; IJkel *et al.*, 2000).

DNA replication and late gene expression

Nineteen late expression factor (*lef*) genes in AcMNPV have been implied in DNA replication and late gene expression (Kool *et al.*, 1995; Lu & Miller, 1995). They were all required for late and very late gene expression. Six of these (*lef-1*, *lef-2*, *lef-3*, *dnapol*, *helicase*, *ie-1*) are essential for DNA replication, whereas others are involved in transcription (*ie-2*, *lef-4*, *lef-5*, *lef-8*, *lef-9*) (Guarino *et al.*, 1998) or in inhibition of apoptosis (such *p35* and *iap*'s) (Clem & Miller, 1994). *In silico* analysis indicated that the genome of HaSNPV contains homologues of 16 of the above AcMNPV *lef* genes and lacks *ie-2*, *p35* and *lef-12* (Table 6-4). The latter genes are also absent in LdMNPV, SeMNPV and XcGV suggesting that they occur only in the group I NPVs. HaSNPV also has a homologue (Ha8) to the first exon of a spliced transcript from Ac141 (*ie-0*). This transcript also includes Ac147 located 4 kb downstream of *ie-0* (Chisholm & Henner, 1988). In contrast, in SeMNPV this exon encoded by Se138 is not functional (Van Strien *et al.*, 2000).

The percentage identity of HaSNPV *lef-8* (Ha38) and *lef-9* (Ha55) with AcMNPV *lef-8* and *lef-9*, encoding subunits of the RNA polymerase complex (Guarino *et al.*, 1998), was the highest with about 65% among the *lef*s, whereas HaSNPV *lef-3* (Ha65) and AcMNPV *lef-3* shared only 27% of their amino acids. HaSNPV LEF3 has a low degree of homology with other NPVs as well (Table 6-2) and a *lef-3* gene is not assigned in XcGV (Hayakawa *et al.*, 1999). It has been suggested that LEF3 is chaperoning other replication factors, such as helicase and LEF2, across the nuclear membrane in infected cells (Wu & Carstens, 1998). Since this membrane is almost eliminated upon infection of cells with GV (Federici, 1999), LEF3 may not be required for GVs to replicate. However, there is a very low degree of homology of *lef-3* to Xc134 and this ORF is also of roughly the same size and has a conserved location in the genome as the other baculovirus *lef* genes. Further

experimentation is required to clarify this assumption. Ha25 shows approximately 36% amino acid identity to Ac25 and Bm16, which encodes a putative DNA binding protein (DBP) (Okano *et al.*, 1999; Mikhailov *et al.*, 1998). An AcMNPV gene involved in the modulation of very late gene expression (*vlf-1*) (Todd *et al.*, 1996) has also been found in HaSNPV (Ha71).

Similar to SeMNPV, LdMNPV and XcGV, HaSNPV also lacks a *p35* homologue (Table 6-4). Instead two members of the *iap* ('inhibitor of apoptosis', Crook *et al.*, 1993) gene family were observed in HaSNPV, *iap-2* (Ha62) and *iap-3* (Ha103). Homologues of *iap* subclasses (1-4) have been found in AcMNPV (Ac27, *iap-1*, and Ac71, *iap-2*), OpMNPV (Op41, *iap-1*, Op74, *iap-2*, Op35, *iap-3*, and ORF106, *iap-4*), SeMNPV (Se88, *iap-2*, and Se110, *iap-3*), LdMNPV (Ld79, *iap-2*, and Ld139, *iap-3*) and XcGV (Xc137, *iap-3*). The HaSNPV *iap-3* gene has high homology to the CpGV *iap* gene, which could functionally complement an AcMNPV *p35* deletion mutant (Crook *et al.*, 1993). OpMNPV *iap-3* can also complement AcMNPV *p35* null mutants (Birnbaum *et al.*, 1994). The function of the *iap-1*, *iap-2* and *iap-4* genes is unknown. Three *iap* gene homologues (*iap-1*, *iap-2* and *iap-3)* were found in *Buzura suppressaria* SNPV through partial DNA sequence analysis (Hu *et al.*, 1998).

HaSNPV lacks genes for enzymatic functions in nucleotide metabolism, such as ribonucleotide reductase (*rr*) and deoxyuridyltriphosphatase (*dUTPase*). The products of *rr* and *dutpase* allow the virus to convert rNTPs into dNTPs to the benefit of virus DNA replication. RR reduces NDPs into dNDPs, and dUTPase converts dUTP into dUMP thereby excluding dUTP from incorporation into DNA and providing dUMP as a precursor for dTTP. *Rr* and *dutpase* are present in SeMNPV (IJkel *et al.*, 1999), OpMNPV (Ahrens *et al.*, 1997) and LdMNPV (Kuzio *et al.*, 1999), but are absent in AcMNPV and BmNPV, and also in XcGV. The latter virus contained a DNA ligase (Xc141) which appeared to be absent in NPVs except LdMNPV.

Genes with auxiliary functions

Baculovirus auxiliary genes are not essential for virus replication *per se*, but are important for example for interaction with the insect host (O'Reilly, 1997). HaSNPV has a very similar set of auxiliary gene as SeMNPV, encoding for example *chitinase* (*chitA*, Ha41), *cathepsin* (*v-cath*, Ha56) and *ecdysteroid UDP-glucosyltransferase* (*egt*, Ha126) (IJkel *et al.*, 1999). These genes are quite well conserved with 66%, 47% and 49% amino acid identity on average, respectively, whereas the fibroblast growth factor (*fgf*, Ha113) is poorly conserved among baculoviruses with on average 28% identity.

HaSNPV lacks a gene for protein tyrosine/serine phosphatase (*ptp*) with dual-specificity (dsPTP) (Tilakaratne *et al.*, 1991; Kim & Weaver, 1993). This protein specifically removes phosphates from both tyrosine and serine/threonine residues (Wishart *et al.*,

Table 6-4. Baculovirus ORFs without homologues in HaSNPV. The AcMNPV ORFs that have no homologue in HaSNPV are shown. ORFs from BmNPV, OpMNPV, LdMNPV and SeMNPV that have no homologue in either AcMNPV or HaSNPV are also shown.

AcMNPV[1]	1 *ptp-1*^{Bm,Op}	20^{Bm}	48 *etm*^{Op}	86 *pnk/pnl*	121	140
	3 *ctl*^{Op,Ld}	27 *iap-1*^{Bm,Op}	49 *pcna*^{Op}	87^{Bm,Op}	122^{Bm,Op}	149^{Bm}
	4^{Bm,Op,Ld}	30^{Bm,Op}	58^{Bm}	91^{Bm,Op}	123 *pk-2*^{Bm}	151 *ie-2*^{Bm,Op}
	5^{Bm,Op}	33^{Se}	70 *hcf-1*	97	124^{Bm,Op}	152
	7 *orf603*	39 *p43*^{Bm}	72^{Bm,Op}	112^{Bm,Ld}	125 *lef-7*^{Bm,Op}	153 *pe38*^{Bm,Op}
	11^{Bm,Op,Ld}	42 *gta*^{Bm,Op}	73^{Bm,Op}	113	128 *gp64*^{Bm,Op}	154^{Bm}
	12^{Bm,Ld}	44^{Bm,Op,Se}	79^{Bm,Op}	116	132^{Bm,Op}	
	16^{Bm,Op}	45^{Bm}	84	118	134^{Se}	
	18$^{Bm,\ Op,\ Ld,\ Se}$	47^{Bm,Op}	85^{Op}	120^{Bm,Op,Ld,Se}	135 *p35*^{Bm,Op}	
BmNPV	111					
OpMNPV	4	28	37	106 *iap-4*^{Ld}	118	147 Opep32
	5	33	68	110^{Ld}	135	148 Opep25
	9^{se}	36	98	113	143 *hrf-1*^{Ld}	149 *p8.9*
LdMNPV	4	11	31	69	127^{Se}	141^{Se}
	5	12	34	77	132	142^{Se}
	6	13	49	111^{Se}	133	144^{Se}
	7 *g22*	22 *ligase*	50 *helic-2*	120	134	152
	8	24	52	121	135	160 *vef-2*
	9	25	59	124^{Se}	137a^{Se}	163
	10	28	65 *vef-1*	126	138	
SeMNPV	5	21	31	83	117	
	17	22	39	85	121	
	18	23	40	86	122	
	20	24	44	116		

1 Superscript means: present in the respective baculovirus

1995). The absence of a *ptp* gene homologue in HaSNPV is striking, since such a gene is present in all NPV genomes sequenced so-far and is thought to be involved in the regulation of the phosphorylation status of viral and host proteins during infection.

Duplicated ORF *bro* genes

A common characteristic of baculovirus genomes is the presence of a group of related genes, the so-named 'baculovirus repeat orfs' or *bro* genes. Five homologues of AcMNPV ORF2 (Ac2) are present in BmNPV (Gomi *et al.*, 1999). In LdMNPV, SeMNPV and XcGV sixteen, one and five *bro*-related genes are found, respectively (Kuzio *et al.*, 1999;

IJkel *et al.*, 1999; Hayakawa *et al.*, 1999). In OpMNPV a truncated version and two smaller *bro*-related ORFs are present (Ahrens *et al.*, 1997). Three *bro*-related genes were identified in HaSNPV, named *bro-a* (Ha59), *bro-b* (Ha60) and *bro-c* (Ha105). Ha59 is most closely related to Ld150 (*bro-m*), belonging to the group II *bro* family (Kuzio *et al.*, 1999), with 50% amino acid identity. Ha60 also belongs to the group II *bro* genes and shares the highest homology to Ld140 (*bro-l*) and Xc159 (*bro-g*), but has an N-terminal duplication of 183 amino acids. It thus seems unlikely that the Ha59 and Ha60 *bro* genes have a common recent ancestor and therefore might have been spliced in tandem into a *hr* sequence (*hr3* and *hr4*). Ha105 and Xc60 are 66% identical and related to the group III *bro* genes (Kuzio *et al.*, 1999).

HaSNPV ORFs with homologues in a few other baculoviruses

HaSNPV possesses twenty-two ORFs that have no homologues in AcMNPV, BmNPV, OpMNPV, SeMNPV, LdMNPV or XcGV (Table 6-2). Of these, Ha6 is identical to HzORF480 from HzSNPV (Le *et al.*, 1997). In HaSNPV a homologue of the *Leucania separata* NPV (LsNPV) *p13* gene (Ha97) is found. This homologue is, in contrast to the SeMNPV homologue, not C-terminally extended (Wang *et al.*, 1995; IJkel *et al.*, 1999). The two leucine-zipper like structures presented in LsNPV P13 (Wang *et al.*, 1995) are also conserved in Ha97. The function of this ORF in LsNPV as well as in SeMNPV (Se59) and XcGV (Xc48) is unknown.

Three HaSNPV ORFs have a homologue in only one other baculovirus. None of these genes have assigned functions yet. Ha19 has a gene homologue in LdMNPV (Ld26) with an amino acid identity of 40%. This ORF however is rather small, encoding an 11 kDa protein. Ha57, encoding a putative 21 kDa product, has a homologue in XcGV (Xc83) with an identity of 33%. A Se68 homologue is identified in HaSNPV as Ha83 encoding a putative protein of 18.8 kDa but with a low amino acid identity (26%). All ORFs, however, have baculovirus early and or late transcription motifs and may therefore be functional.

HaSNPV ORF100 (Ha100) was found to encode a putative poly-(ADP-ribose) glycohydrolase (*parg*). The homology with *Drosophila melanogaster* (24% identity) and *Homo sapiens* (23% identity) was found in the C-terminal portion of the putative protein. Homologues of Ha100 were also found in LdMNPV (Ld141) and SeMNPV (Se52), so that their presence appears to be limited to group II NPVs. In eukaryotes this enzyme is involved in the breakdown of polyribose and recruitment of this compound for nuclear functions such as DNA replication and repair (D'Amours *et al.*, 1999). The function of this enzyme in baculovirus group II morphogenesis or pathology is not known, but it is possible that it is involved in similar capacity during the NPV infection process. The baculovirus *parg* gene is much longer than the eukaryotic counterpart and thus may have additional activities.

A few HaSNPV ORFs (Ha1-4, Ha6, Ha9-11, Ha13-15, Ha41, Ha69, Ha71 and Ha126; Table 6-2) have a high degree of amino acid identity (> 90%) to sequences available from HzSNPV (Ma *et al.*, 1993; Cowan *et al.*, 1994; Le *et al.*, 1997). This suggests that the overall homology between HaSNPV and HzSNPV is very high and that they are most likely variants of the same virus species. Sequencing of the HzSNPV genome would reveal information whether this assumption is correct.

Unique HaSNPV ORFs

To date, twenty ORFs in the HaSNPV genome are unique to this virus and also do not exhibit significant homology to any other sequences in the GenBank. Most of these ORFs are either very small encoding putative proteins up to 100 amino acids (Ha5, Ha7, Ha17, Ha18, Ha40, Ha45, Ha54, Ha104 and Ha112) or contain no common baculovirus transcription initiation sites for early or late gene expression (Ha102, Ha108 and Ha109). Eight ORFs (Ha29, Ha34, Ha99, Ha107, Ha122, Ha125, Ha134 and Ha135) are larger than 100 amino acids and have early and late baculovirus promoter motifs. Ha34 and Ha107 are of interest as they encode putative proteins of 41.1 kDa and 51.2 kDa, respectively. The possible functions of these ORFs are being investigated. For convenience the ORFs present in the other sequences baculoviruses, AcMNPV, BmNPV, OpMNPV, LdMNPV and SeMNPV are listed in Table 6-4.

The HaSNPV genome organization

The genomic organization, i.e. the order of genes, of HaSNPV has been studied in a comparative manner using GeneParityPlot analysis (Hu *et al.*, 1998). As the gene order between AcMNPV, BmNPV and OpMNPV is basically identical, except for a small number of rearrangements (Ahrens *et al.*, 1997; Hu *et al.*, 1998; Gomi *et al.*, 1999), AcMNPV was taken as a representative example of this group in the analysis (Fig. 6-3A). A comparison was made between the recently sequenced MNPVs, SeMNPV (IJkel *et al.*, 1999) and LdMNPV (Kuzio *et al.*, 1999), and XcGV (Hayakawa *et al.*, 1999) (Fig. 6-3B-D). To obtain maximum alignment in the GeneParityPlot analysis the order of genes had to be reversed for the calculation. By convention the orientation of a circular baculovirus genome is determined by the relative position of two genes, *polyhedrin* at map unit 0 and *p10* approximately at map unit 90 (Vlak & Smith, 1982). In the initial GeneParityPlot analysis the orientation of the HaSNPV genome appeared to be reversed for more than 50% of the ORFs compared to AcMNPV and LdMNPV to obtain maximum alignment as compared to the physical map constructed previously (Chen *et al.*, 2000a). A similar situation exists for SeMNPV (IJkel *et al.*, 1999). The gene organization of HaSNPV is most conserved in the 'central region' of the linearized genome and confirms the supposition of Heldens *et al.* (1998) that this region is conserved in all baculoviruses. The left region of the linearized HaSNPV genome displays a considerable number of gene inversions and translocations as deduced from the GeneParityPlot analyses. The right region showed a high degree of gene scrambling (Figure 3A-D). From these analyses it is

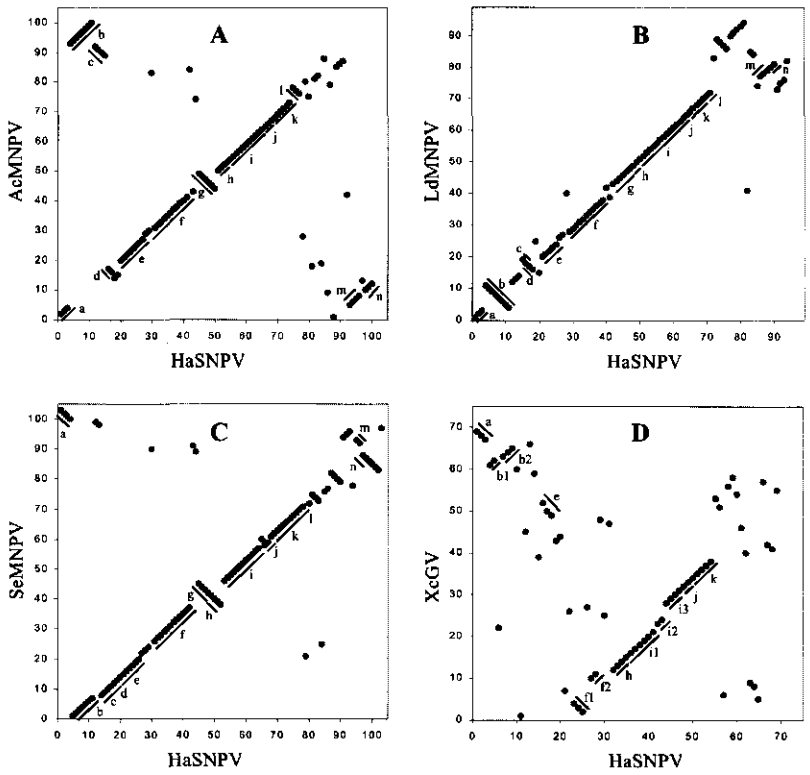


Figure 6-3. GeneParityPlots of HaSNPV versus three other baculoviruses. Graphic representation of the collinearity of baculovirus genomes of AcMNPV (A), LdMNPV (B) and SeMNPV (C) and XcGV (D) obtained by GeneParityPlot analysis (see Materials and Methods). Fourteen putative gene clusters of the HaSNPV genome, which are similar to those of AcMNPV (A), LdMNPV (B) and SeMNPV (C) and were ordered alphabetically (a-n) and underlined. The putative gene clusters are indicated in Table 6-2.

concluded that the organization of HaSNPV is highly characteristic and distinct from those of AcMNPV, SeMNPV, LdMNPV and XcGV.

Comparison of the relative gene order between HaSNPV versus AcMNPV, SeMNPV, LdMNPV and XcGV revealed the presence of certain gene clusters, which are conserved in all baculovirus genomes (Fig. 6-3, Table 6-2). The juxtaposition of ORFs can be used as a phylogenetic marker to study the ancestral relationship of baculoviruses, independent of the evolution of individual genes. These clusters are numbered according to their sequential appearance in the GeneParityPlots. Fourteen clusters conserved in all five baculoviruses have been identified (Fig. 6-3, Table 6-2). In comparison to a previous analysis (IJkel *et al.*, 1999) a small additional cluster, named 12a, (Ac28/Ac29 and their homologues) has been identified. Cluster 12, which was conserved in AcMNPV, LdMNPV and SeMNPV, was interrupted in HaSNPV by a LeseNPV L25 homologue.

Furthermore, the *chitinase* gene (Ha41) has been inserted into cluster 11, whereas Ha40, Ha54 and Ha83 also intervened in this cluster. However, the latter three ORFs are very putative and relatively small genes, in cases (Ha40 and Ha54) without apparent transcription control motifs. One additional cluster has been identified in the GeneParityPlot of HaSNPV versus SeMNPV and LdMNPV encompassing Ha126, Ha128 and Ha129 (Fig. 6-3; cluster n).

The cluster organization of HaSNPV as compared to other baculoviruses (Fig. 6-4) suggests that the genomic organization of HaSNPV is more closely related to SeMNPV and LdMNPV than to Group I NPVs (AcMNPV, BmNPV and OpMNPV) or XcGV. This is in agreement with the phylogenetic analysis of single genes such as *egt*, *lef2*, *dnapol* and *rr* (Chen *et al.*, 1997a, b; Bulach *et al.*, 1999; Chen *et al.*, 1999). When the order of gene clusters is taken to represent the baculovirus genome organization, the common structure of Group II baculoviruses becomes apparent (Fig. 6-4a). Within each group the structural difference is relatively small and predominantly determined by inversions of gene clusters as well as inversions of individual genes (cf. *polyhedrin*). Comparison of the two groups

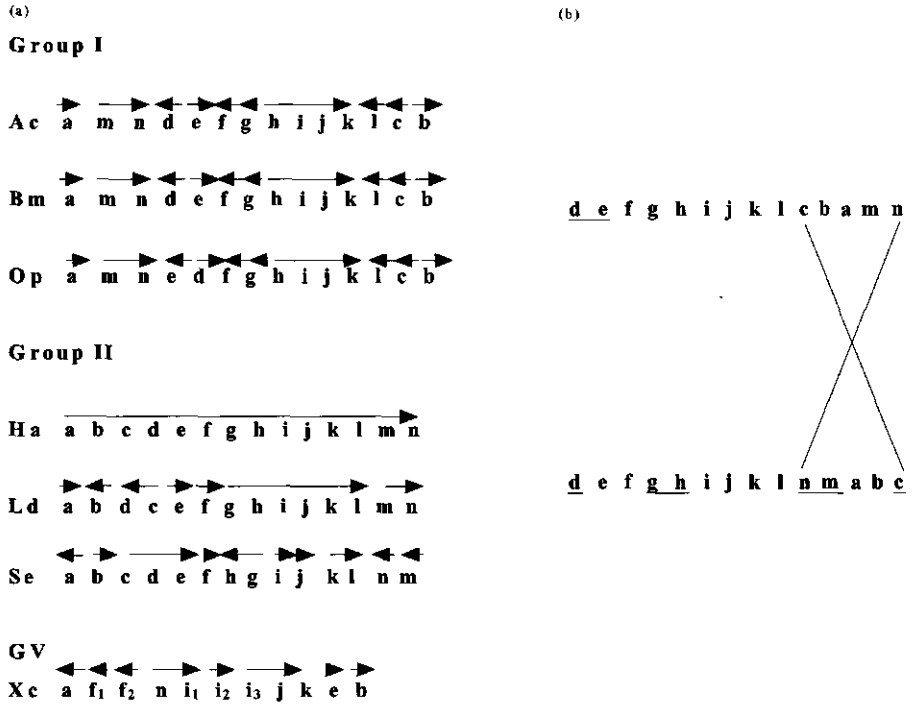


Figure 6-4. Alignment of conserved genome clusters (a-n) of AcMNPV, BmNPV, OpMNPV, SeMNPV, LdMNPV, HaSNPV and XcGV (a) and comparison between Group I and Group II baculoviruses (b). The arrows indicate the orientation of the cluster and the cluster inversions are underlined.

showed extensive genomic translocations in addition to cluster inversions. When the inverted genes remained functional, they could be translocated to other genomic regions. These 'jumping' genes can be used as phylogenetic markers to follow in retrospect baculovirus evolution. A common genome structure for Group I and Group II viruses can be derived showing a major inversion of a genomic segment containing the cluster c-b-a-m-n (Fig. 6-4b).

In conclusion, the sequence of the genome of HaSNPV is distinct from other baculoviruses both in gene content and gene arrangement. Except for three *bro*-related genes and two *iap*-related genes the HaSNPV genome contains 130 unique ORFs, many of which are shared with other NPVs. Based on the percentage identity of gene homologues, on the phylogeny of some particular genes and on the gene arrangement along the HaSNPV genome we conclude that HaSNPV, SeMNPV and LdMNPV must have had a common ancestor. The HaSNPV sequence further confirmed the observation that the part of baculovirus genomes flanking DNA *helicase* is highly conserved possibly as a result of transcriptional or regulatory constraints. By comparing gene clusters a common structural genomic feature is revealed in Group II baculoviruses. Study of the eleven unique putative ORFs (> 100 aa) may provide insight in the determinants specifying the SNPV morphotype. From sequence analysis it is also clear that the SNPV and MNPV morphotype is at best only taxonomic determinant and that SNPVs and MNPVs do not likely represent separate phylogenetic clades.

ACKNOWLEDGEMENTS

This research was supported in part by the Royal Academy of Sciences of the Netherlands and the Chinese Academy of Sciences (98CDP008), the Dutch-Israeli Agricultural Research Program (DIARP) (93/20 and 97/29), and the Natural Science Foundation of China (NSFC). Marjo van Staveren, Marleen Abma-Henkens and Paul Mooijman are thanked for their skillful technical assistance. W.F.J.IJkel. is supported by a fellowship from the Netherlands Foundation for Chemical Sciences (CW) with financial aid from the Netherlands Organization for Scientific Research (NWO). X.C. received a grant from the Royal Academy of Sciences of the Netherlands (KNAW) and a PhD Sandwich fellowship from Wageningen University. Z.H. is a recipient of the Hundreds of Talents Program award (STZ-3-01).

Chapter 7

COMPARATIVE ANALYSIS OF THE COMPLETE GENOME SEQUENCES OF *Helicoverpa zea* AND *H. armigera* SINGLE-NUCLEOCAPSID NUCLEOPOLYHEDROVIRUSES

SUMMARY

The complete nucleotide sequence of *Helicoverpa zea* single nucleocapsid nucleopolyhedrovirus (HzSNPV) has been determined and compared to the nucleotide sequence of *H. armigera* SNPV (Chen *et al.*, 2001). The two genomes are very similar in nucleotide (97% identity) and amino acid (99% identity) sequences and hence largely identical. The coding regions are much more conserved than the non-coding regions. The 65 open reading frames (ORFs) present in all eight baculoviruses sequenced so far are much more conserved than other ORFs in these genomes. HzSNPV has four additional small ORFs compared to HaSNPV; one of them (Hz42) is in a correct transcriptional context. The major differences between HzSNPV and HaSNPV are found in the sequence and organization of the homologous regions (*hrs*) and in the baculovirus repeat orfs (*bro* genes). The sequence homology between the HzSNPV and HaSNPV *hrs* ranges from 90% (*hr*1) to almost 100% (*hr*3) and in the presence / absence of one or more type A and / or B repeats. The HzSNPV *bro* genes differ significantly from those in HaSNPV and may have been acquired independently in the evolutionary past. The sequence data strongly suggest that HzSNPV and HaSNPV are variants of the same virus species, which is supported by their physical and biological properties.

This chapter has been published as

Chen, X., Zhang, W-J., Wong, J., Chun, G., Lu, A., McCutchen, B. F., Presnail, J. K., Herrmann, R., Dolan, M., Tingey, S., Hu, Z. and J. M. Vlak. 2001, Comparative analysis of the complete genome sequences of *Helicoverpa zea* and *H. armigera* single-nucleocapsid nucleopolyhedrovirus. *Journal of General Virology*. Accepted.

Chapter 7

INTRODUCTION

The *Baculoviridae* are a family of viruses with circular, covalently closed, double-stranded DNA genomes ranging in size from 100-180 kb. The rod-shaped virions are occluded into large proteinaceous capsules or occlusion bodies. Two genera have been established: *Nucleopolyhedrovirus* (NPV) and *Granulovirus* (GV), based on particular occlusion body morphology with single (GV) and multiple (NPV) virions occluded in granules and polyhedra, respectively. NPVs are designated single (S) or multiple (M) depending on the number of nucleocapsids packaged in a virion (Van Regenmortel *et al.*, 2000).

Genome sequence analysis has revealed distinctive features of baculoviruses and the extent of their diversity. At present, the nucleotide sequences of six MNPVs, notably *Autographa californica* (Ac) MNPV (Ayres *et al.*, 1994), *Bombyx mori* (Bm) NPV (Gomi *et al.*, 1999), *Orgyia pseudotsugata* (Op) MNPV (Ahrens *et al.*, 1997), *Lymantria dispar* (Ld) MNPV (Kuzio *et al.*, 1999) and *Spodoptera exigua* (Se) MNPV (IJkel *et al.*, 1999), one SNPV, *Helicoverpa armigera* (Ha) SNPV (Chen *et al.*, 2001), and two granuloviruses, *Xestia c-nigrum* (Xc) GV (Hayakawa *et al.*, 1999) and *Plutella xylostella* (Px) GV (Hashimoto *et al.*, 2000), have been determined. The data also showed the genome size ranged from 100,999 bp for PxGV to 178,733 bp for XcGV. Sixty-five ORFs are conserved in all sequenced baculovirus genomes to date (Chen *et al.*, 2001; Hayakawa *et al.*, 2000).

A number of NPVs have been isolated worldwide from insect species belonging to the genus *Helicoverpa* (Lepidoptera: Noctuidae), which include agricultural pest insects such as *Helicoverpa zea*, *H. virescens*, *H. armigera* and *H. punctigera* (Gettig & McCarthy 1982). Those species are major global pests attacking at least 30 different food and fiber crops and in many instances are resistant to chemical insecticides (Ignoffo, 1973; Cunningham, 1998). These NPVs can be divided morphologically into two groups, MNPVs and SNPVs (Blissard *et al.*, 2000). HzSNPV and HaSNPV have been developed as commercial pesticides to control the host pest. HzSNPV was registered as one of the first commercial baculovirus pesticides (Viron-H, Biocontrol-VHZ, Elcar^{TM}) in the 1970s and has been extensively used to control the cotton bollworm in the USA (Shieh, 1989; Cunningham, 1998). HaSNPV, isolated in 1975 in Hubei province, People's Republic of China, has been successfully used in China for over 20 years to control *H. armigera* in cotton and vegetable crops in an area of about 100,000 hectares (Zhang, 1994).

Previous studies have indicated that HzSNPV and HaSNPV are more closely related to each other than to *Helicoverpa spp.* MNPVs (Gettig & McCarthy, 1982; Sun, 1994). The restriction enzyme patterns and even the physical maps of HzSNPV and HaSNPV were quite similar (Knell & Summers, 1984; Chen *et al.*, 2000a). A few genes present in these

two viruses, notably polyhedrin (Chen *et al.*, 1997b), the ecdysteroid UDP glucosyltransferase (*egt*) (Chen *et al.*, 1997a), *hoar* (Le *et al.*, 1997), DNA polymerase (Bulach *et al.*, 1999) share a very high degree of amino acid identity. Recently HaSNPV genome sequence analysis has indicated that HaSNPV ORFs have over 97% amino acid identity on average to reported partial HzSNPV sequences in data bases (Chen *et al.*, 2001). This suggests that the overall homology between HaSNPV and HzSNPV is very high and appear to be variants of the same virus species. This is corroborated by the observation that these two viruses share the same heliothine host range, although they differ in biological activity against individual heliothine species (Hughes *et al.*, 1983). Sequencing of the HzSNPV genome would further substantiate these hypotheses.

In this chapter we describe the complete nucleotide sequence and organization of the *Helicoverpa zea* SNPV genome and make a comparison with the HaSNPV genome published recently (Chen *et al.*, 2001). The major difference occurs in the sequence and organization of the homologous repeat regions. Furthermore, HzSNPV and HaSNV are characterized by the presence of different 'baculovirus repeat orf' (*bro*) genes. This is the first report of a full sequence comparison of two baculovirus variants.

MATERIALS AND METHODS

Insect and virus

H. virescens larvae were obtained from the insectary at Stine-Haskell Research Center (DuPont, Newark, DE). Insects were reared on a wheat germ/soy flour-based diet (BioServe) in 8-oz plastic cups at 27 °C and 40% humidity and under a 16:8 light:dark cycle.

A plaque purified isolate of HzSNPV strain Elcar (Ignoffo, 1965a) was obtained from the USDA at Brownsville, Texas and polyhedra were propagated in *H. virescens* larvae and purified by sucrose gradient centrifugation (O'Reilly *et al.*, 1994).

HzSNPV DNA isolation, cloning, and sequence determination

HzSNPV genomic DNA was isolated from purified occlusion bodies by solubilization in an alkaline solution and viral DNA prepared as described in O'Reilly *et al.*(1994). Size selected viral DNA (2–4 kb) was cloned into the *Sma*I site of M13mp18 (Novagen, Madison, WI) following nebulization of 50 µg of HzSNPV genomic DNA in the Nebulizer 4207 (Inhalation Plastics Inc., Chicago, IL). Fragments were treated with T4 DNA polymerase and the Large Fragment of T4 DNA polymerase (Klenow) according to the manufacturer's protocol to repair the ends prior to cloning. Ligation products were transformed into JM101 cells and ssDNA for sequencing was prepared from over 2000 clones using standard protocols (Sambrook *et al.*, 1989).

Sequencing reactions were performed using the ABI PRISM Big Dye Terminator Cycle Sequencing Ready reaction kit with FS AmpliTaq DNA polymerase (Perkin Elmer) and analyzed on an ABI Model 377 Automated DNA Sequencer. The combined sequence generated from these clones represented 6 fold genomic coverage. Additional sequence for confirmation of ambiguous regions and for filling in gaps in the assembled sequence was obtained from sequencing of PCR products derived from customized oligonucleotide primers.

Assembly of the complete sequence of the HzSNPV genome was accomplished using the PHRAP program (Ewing *et al.*, 1998; Ewing & Green, 1998). The PREGAP4 interface was used to store PHRAP-assembled data in the GAP4 assembly database (Bonfield *et al.*, 1995) and subsequently used for editing and sequence finishing using the GAP4. Consensus calculations with a quality cutoff value of 40 were performed from within GAP4 using a probabilistic consensus algorithm based on expected error rates output by PHRED.

DNA sequence analysis.

Genomic DNA composition, structure, repeats, and restriction enzyme pattern were analyzed with the University of Wisconsin Genetics Computer Group programs (Devereux *et al.*, 1984) and DNASTAR™ (Lasergene). Open reading frames (ORFs) encoding more than 50 amino acids (150 bp) were considered to be protein encoding and hence designated putative genes. One hundred and thirty-five ORFs (out of 139) were checked for maximal alignment with known baculovirus gene homologues extracted from GenBank. The overlap between any two ORFs with known baculovirus homologues was set to a maximum of 25 amino acids; otherwise the largest ORF was selected.

DNA and protein comparisons with entries in the sequence databases were performed with FASTA and BLAST programs (Pearson, 1990; Altschul *et al.*, 1990). Multiple sequence alignments were performed with the GCG PileUp and Gap computer programs (Genetics Computer Group Inc., Madison, Wisconsin, version 10.0) with gap creation and extension penalty set to 9 and 2, respectively (Devereux *et al.*, 1984). Percent identity indicates the percent identical residues between two complete sequences. The GENESCAN program was used for gene predictions. Motif searches were done against the Prosite release 14 database (Fabian *et al.*, 1997; Bairoch *et al.*, 1997). Prediction of transmembrane domains was accomplished with SignalP and PHD software (Nielsen *et al.*, 1997; Rost & Sander, 1993). The DOTTER program was used to identify and classify repeat families and MITEs (Miniature Inverted Repeat Transposable Element).

Nucleotide sequence accession number

The HzSNPV genome sequence has been deposited in GenBank under accession no. AF334030

RESULTS AND DISCUSSION

Nucleotide sequence analysis of the HzSNPV genome

The HzSNPV genome was assembled into a contiguous sequence of 130,869 bp with a G+C content of 39.1% (Table 7-1, Fig. 7-1). This size and G+C content are very similar to the HaSNPV genome of 131,403 bp with a G+C content of 39.1% (Chen *et al.*, 2001). The size of the HzSNPV genome is about 12 kb larger than a previous estimate of 119 kb based on restriction enzyme analysis and physical mapping of DNA fragments (Knell & Summers, 1984). According to the adopted convention (Vlak and Smith, 1982),

Table 7-1 Characteristics of HzSNPV genome

Characteristic	HzSNPV	HaSNPV	Diff. Hz/HaSNPV
Size (kb)	130869	131403	-534
G+C content (mol%)	39.1	39.1	
Total ORFs	139	135	4
Unique ORFs	4		4
Number of *hr* regions	5	5	
Coding sequence (bp)	114031	114070	-39
Intergenic sequence (bp)	9261	9305	-44
hr regions (bp)	7577	8028	-451
Total amino acids encoded by ORFs	38224	38119	105
Number of overlaping ORFs	24	22	2

polyhedrin was designated as the first gene and the adenine residue at the translational initiation codon of the polyhedrin gene as the start point of the circular HzSNPV genome (Fig. 7-1). The derived restriction map is in general agreement with a previously constructed map for HzSNPV (Knell & Summers, 1982), but different from the map based on Hz-15, a plaque-purified strain from HzSNPV-Elcar (Corsaro & Fraser, 1987).

The contiguous sequence of the entire HzSNPV genome was analyzed to identify the potential ORFs by using DNAStar program version 4.1. The potential genes in HzSNPV genome were selected by the following criteria: (i) ORFs larger than 50 amino acids with initiating methionine codons (ATG); (ii). ORFs overlapping less than 25 aa with the adjacent ORFs, except those that the deduced amino acid sequences showed significant homology to ORFs of other baculoviruses or of cellular proteins. A total of 139 putative ORFs and 5 homologous regions (*hrs*) thus identified are illustrated in Figure 7-1 and Table 7-2. On the basis of the above criteria HzSNPV has 4 additional ORFs (Hz26, Hz42, Hz62 and Hz79) than HaSNPV. The coding sequence (11,4031-bp versus 114070-bp) and the intergenic regions (9261-bp versus 9305-bp) are very similar in size in these two viruses. The genome size difference between their genomes (130,896-bp versa

Table 7-2. Listing of potential ORFs in HzSNPV

		HzSNPV					HaSNPV			AcMNPV	
ORF	Name	lef	dir	right	nt	aa	ORF	ID(nt) %	ID(aa) %	ORF	ID (%) %
1	*polyhedrin*	1	>	741	741	246	1	99.3	99.6	8	85
2	*orf1629*	738	>	1979	1242	413	2	99.4	99.3	9	36
3	*pk-1*	1994	>	2797	804	267	3	99.4	100.0	10	41
4	*hoar* *	2922	<	5192	2271	756	4	99.9	100.0	Se4	27
5		5388	>	5567	180	59	5	97.8	96.7		
6		5738	>	6595	858	285	6	99.4	100.0		
7		6807	<	6962	156	51	7	98.7	86.5		
8	*ie-0*	6950	>	7807	858	285	8	99.5	99.3	141	31
9	*p49*	7824	>	9230	1407	468	9	99.4	99.8	142	52
10	*odv-e18*	9241	>	9486	246	81	10	99.4	100.0	143	62
11	*odv-ec27*	9501	>	10355	855	284	11	99.9	100.0	144	52
12		10399	>	10677	279	92	12	100.0	100.0	145	50
13		10704	<	11315	612	203	13	100.0	100.0	146	32
14	*ie-1*	11357	>	13324	1968	655	14	99.8	100.0	147	30
15	*odv-e56*	13377	<	14441	1065	354	15	99.6	99.7	148	49
16	*me53*	14591	>	15670	1080	359	16/17	99.0	98.2	139	26
17		15673	>	15840	168	55	18	100.0	100.0		
18		15893	<	16174	282	93	19	97.2	93.6	Ld26	39
19	*p74*	16195	>	18261	2067	688	20	98.2	98.4	138	53
20	*p10*	18314	<	18577	264	87	21	98.9	98.9	137	26
21	*p26*	18660	<	19463	804	267	22	98.6	98.9	136	37
22		19576	>	19779	204	67	23	99.5	100.0	29	32
23	*lef-6*	19855	<	20418	564	187	24	99.3	99.5	28	30
24	*dbp*	20432	<	21403	972	323	25	99.8	100.0	25	36
25		21547	>	22023	477	158	26	98.3	98.5	26	35
	hr1										
26		23950	>	24102	153	50					
27		24045	<	24812	768	255	27	99.1	98.8	34	31
28	*ubiquitin*	24652	>	24903	252	83	28	99.6	99.8	35	73
29		24967	>	25473	507	168	29	99.4	99.4		
30	*Lsel25*	25492	>	26064	573	190	30	98.8	97.9	Ls25	32
31	*39K / pp31*	26128	<	27063	936	311	31	99.7	99.7	36	36
32	*lef-11*	27029	<	27412	384	127	32	100.0	100.0	37	35
33		27381	<	28097	717	238	33	98.5	99.1	38	54
34		28328	>	29407	1080	359	34	99.5	99.2		
35	*p47*	29475	<	30713	1239	412	35	99.0	98.5	40	56

		HzSNPV					HaSNPV			AcMNPV	
ORF	Name	lef	dir	right	nt	aa	ORF	ID (nt) %	ID(aa) %	ORF	ID(aa) %
36		30786	>	31457	672	223	36	99.3	100.0	41	26
37		31543	>	31785	243	80	37	99.2	100.0	43	25
38	*lef-8*	31782	<	34487	2706	901	38	99.2	99.4	50	64
39		34540	>	35118	579	192	39	99.3	99.0	51	25
40		35259	>	35411	153	50	40	95.4	90.2		
41	*chitinase*	35419	<	37146	1728	575	41	99.3	99.1	126	67
42		37303	<	37509	207	68					
43		37654	<	38199	546	181	42	99.6	99.0	52	29
44		38315	>	38725	411	136	43	100.0	100.0	53	42
45		38732	<	39868	1137	378	44	99.7	99.5	Se107	30
46		39876	<	40103	228	75	45	100.0	100.0		
47	*lef-10*	40063	>	40278	216	71	46	100.0	100.0	53a	43
48	*vp1054*	40151	>	41206	1056	351	47	99.3	99.1	54	44
49		41326	>	41532	207	68	48	100.0	100.0	55	40
50		41533	>	41727	195	64	49	100.0	100.0	56	26
51		42007	>	42498	492	163	50	99.6	98.1	57	42
52		42577	<	43044	468	155	51	98.7	97.4	59	39
53		43056	<	43322	267	88	52	100.0	100.0	60	43
54	*fp*	43534	<	44187	654	217	53	100.0	100.0	61	62
55		44359	>	44544	186	61	54	99.5	98.4		
56	*lef-9*	44670	>	46229	1560	519	55	99.5	99.6	62	65
57	*cathepsin*	46313	<	47416	1104	367	56	99.3	99.5	127	47
58		47457	<	48044	588	195	57	99.0	98.5	Xc83	33
59	*gp37*	48115	<	48954	840	279	58	98.1	98.2	64	57
	hr2		<								
60	*bro-a*	50494	>	51129	636	211	59		39.0	2	21
61	*bro-b*	51240	>	52298	1059	352	60		71.0	2	20
	hr3										
62		52781	<	52960	180	59					
63	*he65*	53027	>	53737	711	236	61	99.6	99.2	105	29
64	*iap-2*	53814	<	54566	753	250	62	98.9	99.6	71	34
65		54614	<	55438	825	274	63	98.4	98.9	69	42
66		55407	<	55808	402	133	64	99.3	99.3	68	42
67	*lef-3*	55828	>	56967	1140	379	65	99.0	99.2	67	27
68		57074	<	59431	2358	785	66	99.2	98.9	66	28
69	*DNA pol*	59462	>	62524	3603	1020	67	99.5	99.5	65	47

	HzSNPV						HaSNPV		AcMNPV	
ORF	Name	lef	dir	right	nt	aa	ORF	ID(nt) %	ID(aa) %	ORF	ID(aa) %
70		62601	<	63059	459	152	68	99.4	98.7	74	26
71	*hzORF384*	63125	<	63508	384	127	69	99.5	100.0	75	24
72		63514	<	63771	258	85	70	99.6	100.0	76	43
73	*vlf-1*	63812	<	65056	1245	414	71	99.5	99.5	77	70
74		65069	<	65401	333	110	72	99.1	99.1	78	44
75	*gp41*	65470	<	66438	969	322	73	99.6	100.0	80	58
76		66368	<	67093	726	241	74	99.7	100.0	81	54
77		66966	<	67643	678	225	75	99.9	100.0	82	31
78	*vp91capsid*	67573	>	70023	2451	816	76	99.1	98.6	83	43
79		70026	<	70202	177	58					
80	*cg30*	70168	<	71019	852	283	77	99.1	98.6	88	31
81	*vp39capsid*	71108	<	71989	882	293	78	99.3	99.7	89	45
82	*lef-4*	71988	>	73373	1386	461	79	99.4	99.8	90	46
83		73426	<	74190	765	254	80	99.4	100.0	92	55
84		74192	>	74680	489	162	81	99.6	100.0	93	54
85	*odv-e25*	74726	>	75418	693	230	82	99.3	100.0	94	44
86		75450	<	75947	498	165	83	99.0	98.8	Se68	26
87	*helicase*	75966	<	79727	3762	1253	84	99.4	99.7	95	44
88		79684	>	80205	522	173	85	99.2	98.9	96	49
89		80264	<	81229	966	321	86	99.5	99.4	98	46
90	*lef-5*	81125	>	82072	948	315	87	99.0	99.1	99	52
91	*p6.9*	82066	<	82395	330	109	88	100.0	100.0	100	44
92		82460	<	83569	1110	369	89	99.6	99.7	101	43
93		83615	<	83983	369	122	90	99.5	100.0	102	26
94		83983	<	85116	1134	377	91	99.8	100.0	103	51
95	*vp80capsid*	85211	>	87028	1818	605	92	99.8	99.5	104	23
96		87025	>	87201	177	58	93	98.9	98.3	110	29
97		87216	>	88301	1086	361	94	99.6	100.0	109	58
98		88346	>	88630	285	94	95	99.0	98.9	108	41
99	*odv-e66*	88697	<	90715	2019	672	96	99.5	99.7	46	43
100	*p13* ‡	90736	<	91566	831	276	97	99.4	98.9	Se56	60
	hr4										
101		93744	>	94343	600	199	98	99.3	99.0	115	40
102		94347	>	94703	357	118	99	97.8	95.0		
103		94798	>	96324	1527	508	100	98.8	98.8	Se52	24
104		96403	>	97164	762	253	101	99.1	99.6	106/107	47/32

	HzSNPV						HaSNPV		AcMNPV	
ORF	Name	lef	dir	right	nt	aa	ORF	ID(nt) %	ID(aa) %	ORF	ID(aa) %
105		97179	>	97511	333	110	102	99.7	100.0		
106	*iap-3*	97570	<	98376	807	268	103	99.1	98.1	Se110	38
107		98373	<	98528	156	51	104	99.4	98.1		
108	*bro-c*	98632	<	100137	1506	501	105	99.3	99.4	Ld71	51
109	*sod*	100305	>	100784	480	159	106	99.0	98.1	31	74
110		100791	>	102164	1374	457	107	99.0	99.6		
111		102218	<	102796	579	192	108	99.7	99.0		
112		102965	>	103312	348	115	109	99.4	99.1		
113		103290	>	103589	300	99	110	99.3	100.0	117	33
114		103657	>	105243	1587	528	111	99.1	99.1	119	48
115		105240	>	105476	237	78	112	100.0	100.0		
116	*fgf*	105499	<	106404	906	301	113	100.0	100.0	32	29
117	*alk-exo*	106531	<	107817	1287	428	114	98.8	98.6	133	44
118		107837	<	108226	390	129	115	100.0	100.0	19	28
	hr5										
119		109615	<	110541	927	308	115a	100.0	100.0		
120		110742	>	110957	216	71	116	100.0	100.0	111	36
121	*lef-2*	111073	<	111789	717	238	117	99.2	99.7	6	43
122	*p24capsid*	112151	>	112897	747	248	118	99.5	99.2	129	37
123	*gp16*	112959	>	113243	285	94	119	100.0	100.0	130	25
124	*calyx*	113295	>	114317	1023	340	120	98.9	99.7	131	36
125		114396	>	114860	465	154	121	100.0	100.0	63	29
126		114991	>	115581	591	196	122	99.3	98.5		
127	*38.7kd*	115625	<	116803	1179	392	123	98.8	98.0	13	27
128	*lef-1*	116805	<	117542	738	245	124	99.3	99.6	14	39
129		117517	<	117951	435	144	125	99.8	99.3		
130	*egt*	118096	>	119643	1548	515	126	99.6	99.4	15	47
131		119843	>	120421	579	192	127	99.8	100.0		
132		120372	>	121172	801	266	128	98.9	97.5	17	24
133		121255	<	124098	2844	947	129	98.8	98.1	Se30	29
134	*pkip-1*	124439	>	124948	510	169	130	99.6	97.1	24	23
135	*arif-1*	125015	<	125812	798	265	131	99.9	100.0	21	29
136		126074	>	127225	1152	383	132	100.0	100.0	22	60
137		127266	<	129299	2034	677	133	100.0	100.0	23	24
138		129441	<	129983	543	180	134	99.5	96.1		
139		130176	>	130763	588	195	135	97.6	95.4		

conserved motif AATCGTGT (-134- -128 nt) upstream of the start codon ATG. It is possible that Hz42 might have a unique function for HzSNPV.

Homologous regions (*hrs*)

Regions with homologous repeats were first found in AcMNPV (Cochran & Faulkner, 1983) and appear to be present in all baculoviruses. They occur at multiple locations along the genome and serve as putative origins of DNA replication (Kool *et al.*, 1995) and as enhancers of transcription (Guarino & Summers, 1986; Guarino *et al.*, 1986). In AcMNPV *hr*s are characterized by the presence of multiple, often imperfect tandemly repeated palindromic sequences. The HzSNPV genome contains five *hr*s and they have the same genomic location and organization as HaSNPV *hr*s. The latter are unique as compared to other baculovirus *hr*s as they have two types of repeats, type A and type B, with imperfect 40-bp long and 107-bp long repeats, respectively, or truncated versions thereof (Chen *et al.*, 2001). The type A and B repeats are also found in each of the *hr*s in HzSNPV.

Sequence alignment between HzSNPV and HaSNPV *hr*s indicated that these homologous regions have a lower nucleotide identity (95%), and that numerous insertions/deletions have occurred. The most divergent region is *hr*1 with 92% sequence identity, which also contains more than 10 insertions/deletions. The biggest deletion in HzSNPV *hr*1 compared with HaSNPVs is a 271-bp segment, which contains both type A and B repeat, while a 171-bp insertion contains only a type B repeat. The other deletions /insertions, ranging from 1 to 59 bp, did not contain an entire repeat. In most cases they were short to contain an entire repeat. This situation also occurred in the other *hr*s. *Hr*2 with the lowest identity of 90.5% contains a few deletions/insertions of different sizes. No major deletion/insertion was found in *hr*3, the smallest *hr*, but the sequence identity is only 94.4%. As indicated before (Chen *et al.*, 2001), *hr*2 and *hr*3 might have been originally a single *hr* that was split by two *bro* genes. So the *hr*2 – *hr*3 region stands out as the most divergent region in the Hz/HaSNPV genome. *Hr*4 has 95% sequence identity and contains single 132-bp deletion, while the size of other small deletion/insertion ranging from 3-12 bp. The most stable is *hr*5 with 98.6% sequence identity, a single 290-bp deletion and contain type A and B repeats.

Comparison of different HzSNPV isolates indicated that there are four genomic regions with heterogeneity (Corsaro & Fraser, 1987). In HaSNPV six variable regions were identified (X. Chen and Z. Hu, unpublished). Comparison of HzSNPV and HaSNPV suggests that five of these hypervariable regions are located in the *hr*s region. In one case the heterogeneity was located in the *haor* gene in HaSNPV. The same type of variation was detected for several variants in *hr*s of CfMNPV (Arif & Doerfler, 1984), four genotypic variants of SeMNPV (Muñoz *et al*, 1998) and two variants of AgMNPV (Garcia-Maruniak *et al.*, 1996). These baculovirus variants contain different numbers of

repeats suggesting that *hr*s have a high structural plasticity. It is interesting to note that, despite the high degree of structural similarities among *hr*s, the actual sequences are entirely different among baculovirus species. For example, the major difference in genomic organization between AcMNPV and OpMNPV occurred in the homologous repeat regions (Ayres *et al.*, 1994; Ahrens *et al.*, 1997). Since *hr*s are involved in regulating processes (enhancing transcription, DNA replication), they may also be involved in recombination or virulence.

***Bro* genes**

A common characteristic of baculovirus genomes is the presence of a group of related genes, the so-named 'baculovirus repeat orfs' or *bro* genes. Five homologues of AcMNPV ORF2 (Ac2) are present in BmNPV (Gomi *et al.*, 1999). Bro-related genes have also been found in LdMNPV (16), SeMNPV (1) and in XcGV (5).(Kuzio *et al.*, 1999; IJkel *et al.*, 1999; Hayakawa *et al.*, 1999). In OpMNPV a truncated version and two smaller *bro*-related ORFs are present (Ahrens *et al.*, 1997). Three *bro*-related genes were identified in HaSNPV (Chen *et al.*, 2001), but no *bro* was found in PxGV (Hashimoto *et al.*, 2000).

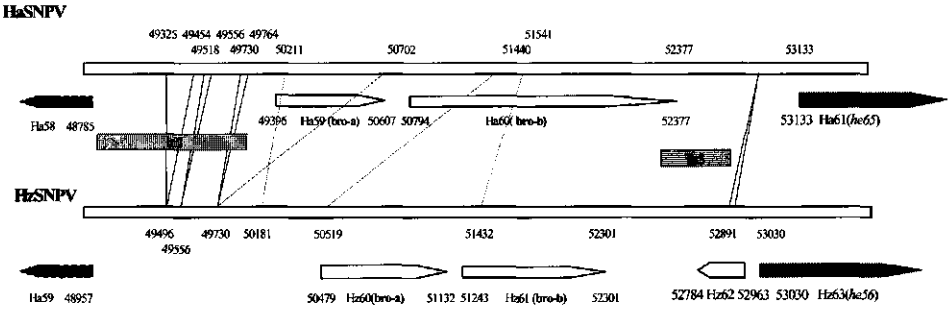


Figure 7-2. Comparison of HzSNPV *hr2-hr3* region with HaSNPV. The gene organization of *hr2-hr3* is presented here. The arrows indicate the ORFs and their direction of transcription. The boxes represent the *hr*s. The major deletions/insertions were located in the HzSNPV/HaSNPV genomes, with the corresponding number above or under the lines. One region (nt 49730-50181-50519 nt) in HzSNPV showed a low homology to two regions in the HaSNPV genome (nt 49764-50211; nt 50702–51440) indicated by the dot-line. The short insertions/deletions are not indicated.

Three *bro* genes, named *bro-a* (Hz60), *bro-b* (Hz61) and *bro-c* (Hz108), were identified in HzSNPV. A similar number of *bro* genes was present in HaSNPV and at the same location. In contrast, however, *bro-a* and *bro-b* have a very low amino acid homology with the corresponding ORFs in HaSNPV. Hz60 is 33 aa smaller than Ha59 with only 39.5% identity. Hz61 shares 71% identity with Ha60, but Ha60 has an N-terminal duplication of 183 amino acids. These additional 183 N-terminal amino acids of Ha60 also showed significant homology with Hz60 (48%), Hz62 (59%) and Ha59 (43.9%) as well as

the C-terminus of Ha60 (52%).

The HzSNPV BROs were compared to other baculovirus BROs by phylogenetic analysis. Based on amino acid sequence homologies, the BROs are divided into two groups: group A (*bro* group) and group B (*bro*-like group). Group A and group B BROs have the same structure but the homologies between them are quite low. Two unrooted parsimonious trees were constructed for the two groups (Fig. 7-3). The baculovirus BROs were divided into four sub-groups, A-I, A-II, B-I and B-II. Group A-I appears to encompass Ac-BRO (AcORF2), five BmNPV BROs, Op116 and the LdMNPV BRO-B, J, K, N, P.. This grouping is in agreement with the BRO group proposed previously (Kuzio *et al.*, 1999). Within this clade, three further subgroups could be distinguished (Fig 7-3a). These BROs have relatively high homology with AcORF2. Hz60, Hz61, Ha59 and Ha60 belong to the Group A-II genes. The evolutionary history of Ha60 appears to be rather complex. Only if Ha60 was separated into two domains (Ha60-N containing 182 N-terminal amino acids and Ha60-C containing the remainder of the protein), it becomes apparent that Ha60-N and Ha60-C have evolved from a different but a recent ancestor. It can also be seen that Hz61 and Ha60-C had a common and a recent ancestor (Fig. 7-3a). Ha59 was more closely related to Ld-BRO-M than to Hz-BRO-A. It is possible that HzSNPV and HaSNPV might have acquired those *bro* genes from different sources. Group B-I contains only the BROs from LdMNPV that once were divided into two subgroups according to their structural domains (Kuzio *et al.*, 1999) (Fig 7-3b). Hz108, Ha105, Xc-BRO-A, Xc-BRO-E, Xc-BRO-F and Op68 all belong to Group B-II and the topology of this part of the tree is well supported by high bootstrap values. It is notable that the BROs from the same virus do not always have a close relationship suggesting that they might have a different evolutionary history. In a particular virus, some might be the result of gene duplication and others might have been picked up from other sources, for example, their insect host or another baculovirus occupying the same ecological niche.

It has been demonstrated that *bro-d* is essential in the BmNPV infection cycle, while *bro-a* and *bro-c* may functionally complement each other (Kang *et al.*, 1999, Zemskov *et al.*, 2000). BRO-A, BRO-C and BRO-D have nucleic acid binding activities and are involved in nucleosome organization that could block cellular replication and /or transcription and switch the host machinery to viral DNA or RNA synthesis. However BRO-B and BRO-E are found in the cytoplasm of infected cells, while BRO-A, BRO-C and BRO-D are present in the nuclei. Deletion of *bro-b* and *bro-e* did not altered BmNPV infectivity (Kang *et al.*, 1999) suggesting that different BROs may have different functions. Since HzSNPV/HaSNPV *bros* are in different phylogenetic groups with BmNPV, it is possible that they have different functions from BmNPV BROs.

Sequence divergence between HzSNPV and HaSNPV

The sequence divergence between homologous genes and the corresponding intergenic

Figure 7-3. Phylogenic trees of baculovirus BROs. To achieve the highest alignment the baculovirus BROs were divided into two groups, *bro* and *bro*–like, and two unrooted parsimonious trees were constructed by using the branch- and-bound algorithm of PAUP3.1. HaSNPV BRO-B was divided into two parts: N terminal part (Ha-BRO-B-N) and C- terminal part (Ha-BRO -C). The details are described in the text. The BROs from HzSNPV and HaSNPV are in bold. Numbers above the line indicate the phylogenetic distance. Numbers in italics indicate the frequency of a given cluster after bootstrap analysis (100 replicates).

regions from HzSNPV and HaSNPV have been investigated in detail and a summary is presented in Table 7-3 and Fig. 7-1 and 7-4. The hypervariable *hrs* regions and *bro-a* and *bro-b* mentioned earlier are not included in this analysis.

A total of 882 nucleotide substitutions and 57 insertions / deletions have been found in HzSNPV genome when comparing the entire nucleotide sequence with that of HaSNPV. The nucleotide substitution frequency on average is 7.15‰. These mutations are distributed along the genome but not randomly (Fig. 7-1). The 9261 bp non-coding sequence accumulated a total of 117 nucleotide-substitutions while 765 nucleotide-substitutions occurred in coding sequences. The substitution frequency is about two times higher in the intergenic sequence (11.7‰) than in the coding sequence (6.8‰). There are

also considerably more deletions/insertions in the noncoding regions than in the ORFs (Table 7-3). In the noncoding sequence, the regions between Hz133 and Hz135, Hz 135 and Hz136, Hz138 and Hz139 were 'hot spots' for alterations. However, there are only 5 deletions/insertions in the 'central part' of the genome, from Hz63 to Hz100 in a total of about 37 kb of sequence. This correlates to the observation that this part of baculovirus genomes is highly conserved, possibly as a result of transcriptional or regulatory constraints (Heldens *et al.*, 1998; Hu *et al.*, 1998; Chen *et al.*, 2001). No insertions/deletions were observed in the 65 conserved baculovirus genes.(Chen *et al.*, 2001). Furthermore, if there is a deletion/insertion in the coding sequence, it always seems to occur at the C-terminus of the protein and in most cases did not cause a frame shift (Table 7-2). Regardless of the coding or non-coding nature of sequences, transitions are dominant where a purine is replaced by a purine or a pyrimidine by a pyrimidine (69.7%). Again these substitutions appear to take place at different rates between coding (71.0%) and non-coding sequences (60.7%). Apparently the coding regions are under a different selection pressure than the non-coding sequences allowing more substitutions in the latter as long as there is compensation to maintain the structure.

Table 7-3. Mutation analysis in HzSNPV /HaSNPV*

	Coding sequence	Intergenic sequence
Number of nucleotides substituted		
Total	765	117
Transition#	544	71
Transvertion##	221	46
Nonsynonymous	228	
Deletions	14	18
Insertions	8	15
Average frequency of substitution(‰)	HzSNPV	HaSNPV
in the genome	7.14	7.15
in the coding sequence	6.69	6.71
in the intergenic sequence	12.69	12.57
Frequency of nonsynonymous substitution	1.99	2.00
Frequency of amino acid substitution	6.01	6.03

* *hr* regions and *bro*-a and *bro-b* were not included in this table
\# one purine nucleotide is replaced by another, or one pyrimidine nucleotide is replaced by another
\# \# one purine nucleotide is replaced by one pyrimidine, or one pyrimidine nucleotide is replaced by purine

The average frequency of nucleotide substitution in coding regions of HzSNPV/HaSNPV is 6.8‰. For individual genes, the substitution frequency ranges from 0 (21 ORFs mentioned earlier) to 45.8‰ for Hz40. Seven nucleotide substitutions, resulting in a 5 amino acid-difference in a 150-bp stretch of coding sequence was found in this ORF, which so far is unique to HaSNPV and HzSNPV. On average, the identity of the proteins involved in DNA replication and regulation of the gene expression and that of the structural proteins are 99.6%, while with auxiliary proteins it is 99.4%, which is higher

than the mean identity of all ORFs (98.3%). The 65 proteins conserved among all baculoviruses have a lower divergence in these related baculoviruses.

The extent of synonymous substitutions also varies among ORFs, which is indicated by the different frequency of synonymous and nonsynonymous divergence in HzSNPV genes (Fig. 7-4). The extent of amino acid and nucleotide sequence divergence is correlated, but the relationship between these is interesting in this case. In genes encoding highly conserved proteins, for example the 65 conserved baculovirus proteins, amino acid sequence identity exceeds nucleotide identity because apparently only silent (synonymous) substitutions are permitted in these genes. However, in less conserved genes or ORFs, nucleotide similarity exceeds protein similarity. This presumably reflects the degeneracy of the genetic code: for example, a single codon with nucleotide

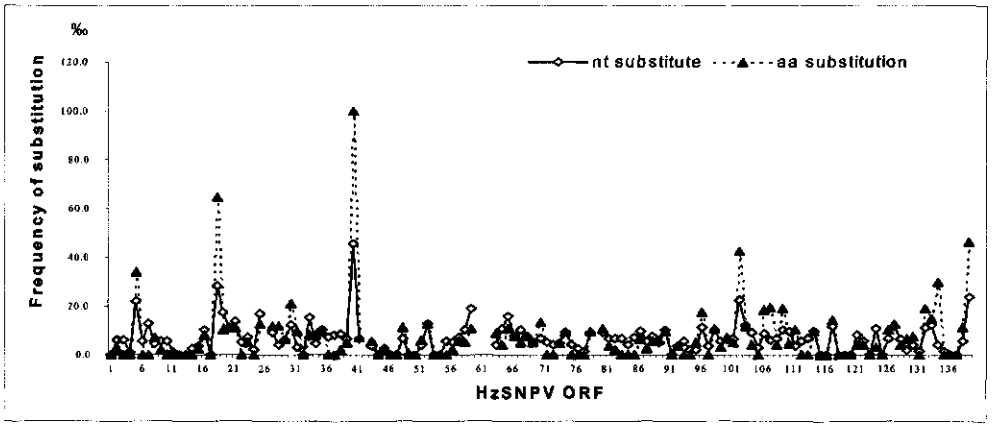


Figure 7-4. Frequency of nucleotide or amino acid substitution in HzSNPV ORFs compared with the corresponding ORFs in HzSNPV. The unique ORFs and Hz60 and Hz61 are not included.

substitutions at positions 2 and 3 (the former resulting in an amino acid replacement, the latter probably silent) exhibits 33% nucleotide sequence identity but 0% amino acid sequence identity. This situation is the same as that observed in rodent genes (363 genes compared between mouse and rat, Wolfe & Sharp, 1993), but markedly different from that seen in a study of bacterial sequence divergence (67 genes compared between *Escherichia coli* and *Salmonella typhimurium*, Sharp, 1991). Bacterial protein sequences are invariably more similar than nucleotide sequences over a range of 76 to 100% amino acid sequence identity. This difference is probably related to a much higher mean ratio of synonymous-to-nonsynonymous divergence in bacterial genes.

CONCLUSION

The complete HzSNPV genome sequence has been analyzed and compared to other baculovirus genomes, particularly that of HaSNPV (Chen *et al.*, 2001). One hundred and

thirty-nine ORFs with larger than 50 amino acids and 5 *hrs* were identified. The genome of HzSNPV is very similar to that of HaSNPV, but distinct from other baculovirus genomes both in gene content and gene arrangement. Except for 4 unique ORFs and two *bro* genes, 133 HzSNPV ORFs have a very high identity (on average 98.9%), to their corresponding ORFs in HaSNPV. Gene translocation was not observed, but nucleotide substitutions and the deletions/insertions with sizes ranging from a few base pairs to 334 bp were found. Sequence alignment of the HzSNPV and HaSNPV genomes indicated that the *hrs* are the most variable, while the 'central region', *i.e.* flanking the *helicase* gene is the most conserved (Heldens *et al.*, 1998). Genes common to all baculoviruses have a much lower degree of divergence than other ORFs. The two *bro* genes, *bro*-a and *bro*-b, might have different origins.

We conclude that HzSNPV and HaSNPV are variants of the same virus species based on genome sequence features and other structural and biological properties, including the morphotype (single nucleocapsid nucleopolyhedrovirus, SNPV), common heliothine host, and similar virulence to the insect hosts (Hamm, 1982; Hughes *et al.,* 1983; Kelly *et al.*, 1980; Williams and Payne, 1984; Sun and Zhang, 1994). *H. virescens* (HeviSNPV) and *H. punctigera* (HepuSNPV) may also belong to the same virus species based on similar physical maps (Gettig and McCarthy, 1982). Different genotypic variants of HzSNPV and HaSNPV have been reported. The polymorphism in the HaSNPV isolate (G4) was at approximately 100 nucleotide locations (0.07%) (Chen *et al.*, 2001). This indicates that the substitution ratio between HaSNPV isolates (genotypes) is much lower when compared to the HzSNPV genome. Taking into account the hypervariable *hrs*, the *bro* genes and 4 unique ORFs, it is clear that HzSNPV and HaSNPV are not as closely related to each other as the genotype isolates of HaSNPV. We hypothesize that HaSNPV and HzSNPV are adapted/adapting to their respective host insects, *H. armigera* and *H. zea*. Most likely the same situation exists with the baculovirus type species, AcMNPV, and with *Anagrapha falcifera* (Anfa) NPV, *Galleria mellonella* (Gm) NPV, *Rachiplusia ou* (Ro) MNPV, *Spodoptera exempta* (Spex) MNPV and *Trichoplusia ni* (Tn) MNPV being genotypic variants, that have diverged into strains with distinctive profiles. The sequence information as presented in this chapter calls for a revision of baculovirus nomenclature (Van Regenmortel *et al.*, 2000) to better reflect the genetic information.

.

ACKNOWLEDGEMENTS

This research was supported in part by the Dutch-Israeli Agricultural Research Program (DIARP) (97/29), X. C. received grants from the National Science Foundation of China and the Royal Academy of Sciences of the Netherlands (KNAW) and a PhD Sandwich Fellowship from Wageningen University.

Chapter 8

GENETIC ENGINEERING OF *Helicoverpa armigera* SINGLE-NUCLEOCAPSID NUCLEOPOLYHEDROVIRUS AS AN IMPROVED PESTICIDE

SUMMARY

The *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus (HaSNPV) has been registered and is commercially produced in China as a bio-pesticide to control the bollworm in cotton. However, the virus has a relatively slow speed of action. To improve its efficacy, recombinant HaSNPVs were generated by deleting the ecdysteroid UDP-glucosyltranserase (*egt*) gene (HaCXW1 and HaLM2) or by inserting the insect-specific toxin gene AaIT in the *egt* locus (HaCXW2) of HaSNPV using conventional recombination strategies in insect cell culture. The various recombinants remained genetically stable when cultured in HzAM1 insect cells. Bioassay data showed a significant reduction in the time required for all HaSNPV recombinants to kill second instar *H. armigera* larvae. The LT_{50} of the *egt* deletion recombinants HaCXW1 and HaLM2 was about 27% faster than that of HaSNPV-wt. The largest reduction in LT_{50} was achieved by inserting the gene for the insect-specific neurotoxin, AaIT, in the *egt* locus giving a reduction in LT_{50} of 32% as compared to wild type HaSNPV. The ability to genetically improve the properties of HaSNPV as a biopesticide provides a further opportunity to develop this virus into a commercially viable product to control the bollworm in China.

This chapter has been published as

Chen, X., Sun, X., Hu, Z., Li, M., O'Reilly, D. R., Zuidema D. and J. M. Vlak. 2000, Genetic engineering of *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus as an improved pesticide. *Journal of Invertebrate Pathology* 76:140-146.

Chapter 8

INTRODUCTION

Baculoviruses comprise a family of double–stranded DNA viruses, which are pathogenic for arthropods, mainly insects. These viruses offer an attractive means to control specific insect pests with minimal adverse effects on the biotic environment. Some baculoviruses have been used successfully in the control of a range of Lepidopteran and Hymenopteran insects (Huber, 1986; Moscardi, 1999). Among the biological features which make these viruses attractive as bio-control agents are their efficacy and specificity. However, baculoviruses have some limitations, which have restricted their widespread commercial use. Their major disadvantage is that for most baculoviruses it typically takes from 4-14 days to kill the insect host. During this time, the insects can still cause serious damage to the crop.

Genetic engineering of the molecularly well-characterized *Autographa californica* nucleopolyhedrovirus (AcMNPV) has led to the improvement of this virus as a pesticide (Possee *et al.*, 1993; Black *et al.*, 1997). The improvement resulted in a reduction of the time required to kill the target pest insect or to prevent it from further feeding. A variety of strategies has been implemented to achieve this ranging from deleting viral genes, which normally function to extend the lifetime of the infected insect, to inserting foreign genes, whose expression in the target insect interferes with some critical aspects of its physiology and results in feeding arrest or death of the insect (O'Reilly and Miller, 1991; Tomalski and Miller, 1991; Stewart *et al.*, 1991; Maeda *et al.*, 1991).

Baculoviruses encode an ecdysteroid UDP-glucosyltransferase (EGT) which catalyses the transfer of a sugar, either glucose or galactose, to ecdysteroids and thereby inactivates the ability of these hormones to induce molting and pupation of the infected host (O'Reilly and Miller, 1989, 1990; O'Reilly *et al.*, 1992; Barrett *et al.*, 1995). Deletion of the *egt* gene of baculoviruses often reduces the time required to kill the target pest insect and also reduces the amount of food consumed by the infected larvae (O'Reilly and Miller, 1991; Slavicek *et al.*, 1999).

The most rapidly acting viruses are those over-expressing insect specific neurotoxin genes (Stewart *et al.*, 1991; Tomalski and Miller, 1991; McCutchen *et al.*, 1991; Maeda *et al.*, 1991; Popham *et al*, 1997; Gershburg *et al.*, 1998). These recombinants secrete the toxin into the insect body and cause neuromuscular effects or acute paralysis. The added benefit is that host feeding is greatly reduced (Black *et al.*, 1997; Hu and Vlak, 1997). Among these toxin genes the *Androctonus australis* inhibitory toxin AaIT has been considered to be the most potent anti-insect specific neurotoxin due to the immediate contraction of larval bodies. Hence this toxin has become a primary tool to engineer baculoviruses for improved insecticidal activity (Stewart *et al.*, 1991; McCutchen *et al.*, 1991)

Since the 1970s the *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus (HearSNPV or HaSNPV) has been used to control the bollworm in cotton in China and has been registered and commercially produced as a biopesticide (Zhang, 1994). This product, however, does not compete favorably with chemical pesticides primarily due to the relatively slow speed at which the virus kills the insect pest. With an interest in isolating (Sun *et al.*, 1998) or generating improved strains of HaSNPV (this paper), we determined whether the viral efficacy could be enhanced through deletion of the *egt* gene and/or expression of the AaIT gene. The homologues of baculovirus *egt* and polyhedrin genes in HaSNPV have been identified and characterized (Chen *et al.*, 1997a, b). A detailed physical map has been constructed recently and a genomic comparison with other baculoviruses has been made (Chen *et al.*, 2000). In this paper we describe the generation of recombinant HaSNPVs, which no longer produce EGT and express the AaIT gene to give these viruses improved insecticidal properties. A green fluorescent protein (GFP) marker was introduced to select the recombinant *in vitro* (Wilson *et al.*, 1997) and to subsequently follow its fate in the environment.

MATERIALS AND METHODS

Insect cells and virus

The *H. zea* cell line Hz-AM1 (McIntosh and Ignoffo, 1983) was used and maintained at 28° C in Grace's medium supplemented with 10% fetal bovine serum. The HaSNPV, which was isolated in Hubei, People's Republic of China, served as the parental wild type (wt) virus and was propagated in Hz-AM1 cells. Viral DNA was isolated from polyhedra as described previously (Sun *et al.*, 1998).

Construction of transfer vectors

The HaSNPV polyhedrin promoter (Chen *et al.*, 1997b) was chosen for the expression of the AaIT toxin and the marker (GFP) gene. The 150 bp polyhedrin promoter was synthesized from the *Hin*dIII-D fragment by PCR with the primers VW246: 5' GC *CTG CAG CCT GAG GG*A TTT CTG TCG TGT TGA and VW247: 5' GC *CTG CAG* AAT TAT GGG ATA TTT GAT TTT TC. *Pst*I sites (italics) were introduced for further cloning. Since the HaSNPV genome does not contain a *Bsu*36I site, such a site was added at the 5' end of the promoter for linearization of the viral DNA (Kitts *et al.*, 1990). The transcription start codon ATG of polyhedrin open reading frame was changed to ATT to allow the exploitation of the ATG of the AaIT insert.

The GFP gene was used as a marker to screen the recombinants. Plasmid pVP92GFP, kindly provided by Dr H. Reiländer, contains the *Aequorea victoria* GFP preceded by the AcMNPV polyhedrin promoter (Reiländer *et al.*, 1996). The GPF cassette was isolated from an agarose gel after excision from the plasmid by *Eco*RV and *Hin*dIII digestion and was inserted into the same sites of pBluescriptIIKS (+) giving pGFP.

The HaSNPV *egt* gene spans the *Hin*dIII-D and *Hin*dIII-I fragments (11.3-12.5 m.u.; Fig. 8-1) (Chen *et al.*, 1997a; Chen *et al.*, 2000). A 2.0-kb *Eco*RI - *Eco*RV fragment from *Hin*dIII-D, which contains the 3' end of the *egt* gene and flanking sequence, was first cloned into pKSPS$^-$ (derived from pBluescriptIIKS (+) by removing the cleavage sites from *Sst*I to *Pst*I in the latter plasmid) to give the intermediate plasmid pHaCXWa. The GFP gene preceded by the AcMNPV polyhedrin promoter was cloned as a *Hin*dIII - *Sma*I fragment from pGFP into the *Hin*dIII-*Eco*RV sites of pHaCXWa, giving intermediate plasmid pHaCXWb. The plasmid pHaCXW0 was then generated by subcloning a 1.7 kb *Pst*I (blunt)-*Hin*dIII fragment from *Hin*dIII-I, which contains 114-bp of the 5' end of the *egt* open reading frame (ORF) and the upstream flanking sequences, into *Xho*I (blunt)-*Hin*dIII cut pHaCXWb. Plasmid pHaCXW0 was cut with *Pst*I and the AcMNPV polyhedrin promoter replaced by the PCR product containing the HaSNPV polyhedrin promoter to generate the transfer vector pHaCXW1 (Fig. 8-1).

The AaIT gene (McCutchen *et al.*, 1991) flanked with *Hin*dIII sites was synthesized with the primers VW248: 5' GC *AAG CTT CTG* CAG GGA TCC GAT CGC ATG AAG and VW249: 5' GC *AAG CTT* GAT CTG AAC TTG TTT ATT GC from the plasmid pAcRH1 (Martens *et al.*, 1995) and was cloned into the *Hin*dIII site of pKSPS$^-$. Then the HaSNPV polyhedrin promoter, as a PCR product, was inserted at the *Pst*I site upstream of the AaIT toxin gene. This AaIT cassette was then inserted into the *Hin*dIII site of pHaCXW1 resulting in the transfer vector pHaCXW2 (Fig. 8-1). Plasmid pHaCXW2 was digested with *Bsu*36I and self-ligated producing a further transfer vector, pHaLM2 (Fig.8-1), which lacks the AaIT gene as well as an intact *egt* gene.

Generation of the recombinant viruses

Hz-AM1 cells (2 x 10^6) were cotransfected with 0.5 μg HaSNPV DNA (either circular as for wt HaSNPV or linearized at the *egt* locus by *Bsu*36I as for HaCXW2) and 5 μg of appropriate transfer vector DNA using Lipofectin™ (GIBCO BRL/life Technologies) as described (King and Possee, 1992). Recombinant plaques were identified either by strong green fluorescence under UV due to GFP where GFP had been inserted, or by the lack of fluorescence when pHaLM2 was used. All plaques were polyhedron-positive as the alteration takes place at the *egt* locus. The putative recombinants were purified by at least three rounds of plaque-purification. Viral DNA isolated from the polyhedra was analyzed by restriction endonuclease analysis and by PCR with the primers VW243: 5' GCG ACG CCA CTA TAA GAT GC and VW245: 5'AAC CGC CGA ATA GCC TGA CC to confirm the allelic replacements.

***In vitro* EGT activity assay**

Hz-AM1 cells (1x10^6/35 mm Petri dish) were infected with wild type (wt) or recombinant HaSNPV at 10 $TCID_{50}$ units/cell. Infection of Sf21 with AcMNPV (wt) or the recombinant AcMNPV-RM2, an *egt*-deletion mutant (Bianchi *et al.*, 2000), served as

controls for the presence and absence of EGT, respectively. Seventy-two hours post infection the infected cell culture supernatant was collected and EGT activity was assayed as previously described (O'Reilly *et al.*, 1992). Radiolabeled ecdysone was separated from ecdysone-sugar conjugates by thin layer chromatography on silica gel plates and radioactivity was detected by autoradiography.

Bioassay

Prior to the bioassay, wild type and recombinant viruses were propagated in *H. armigera* and purified as described by Sun *et al.* (1998). The LC_{50} and ST_{50} of the viruses were determined using second instar *H. armigera* by a modified droplet-feeding bioassay method as described by Hughes *et al.* (1986). To determine the LC_{50} values, larvae were exposed to six concentrations: 1×10^5, 3×10^4, 1×10^4, 3×10^3,1×10^3 and 0 PIBs/ml of each virus (36 insects per dose) and checked for mortality every 12 h. To determine the ST_{50} values, larvae were fed on 10^6 PIBs/ml of each virus, by providing them with 2-5 µl droplets of the PIBs suspended in 5% sucrose and 1 ng/ml FD&C blue No. 1 dye and placed in the center of a Φ60 mm glass Petri dish. Larvae that had ingested the PIB suspension within 30 min, as determined by the blue coloration of their midguts, were transferred to fresh diet. The larvae were maintained at 28°C and monitored approximately every 8 h during the experiment. Bioassays were repeated twice. LC_{50} values were determined using the computer program POLO (Russell *et al.*, 1977) and ST_{50} values were calculated using the ViStat program (version 2.1; Boyce Thompson Institute, Cornell University, Ithaca, New York; Van Beek and Hughes, 1998). LogLC_{50}, ST_{50} values and the slopes of dosage-mortality and time-mortality responses were analyzed by regression analysis and a t-test of pairwise differences between treatments using Genstat (Payne *et al.*, 1993).

RESULTS

Construction of the transfer vectors

The transfer vector pHaCXW1 was constructed by joining a 2.0 kb *Eco*RI - *Eco*RV fragment from *Hin*dIII-D, a GFP cassette and a 1.7 kb *Pst*I-*Hin*dIII fragment from *Hin*dIII-I (Fig. 8-1). In pHaCXW1 the major part of the *egt* open reading frame (ORF) is replaced by the GFP reporter gene under the control of the HaSNPV polyhedrin promoter in the *egt* antisense orientation. Plasmid pHaCXW1 was used as a parental plasmid to generate other transfer vectors, such as pHaCXW2 and pHaLM2 (Fig. 8-1). In the transfer vector pHaLM2 a major part of the *egt* ORF is deleted and there is only a unique *Bsu*36I site at this locus. Plasmid pHaCXW2 was derived from pHaCXW1 by insertion of an AaIT cassette, comprising a signal sequence, the AaIT gene and SV40 terminal sequences flanked by *Hin*dIII sites. This cassette was synthesized from pAcRH1 (Bianchi *et al.*, 2000) via PCR. The inserts were checked by sequencing (data not shown). For over-expression of the AaIT, the polyhedrin promoter of HaSNPV was used.

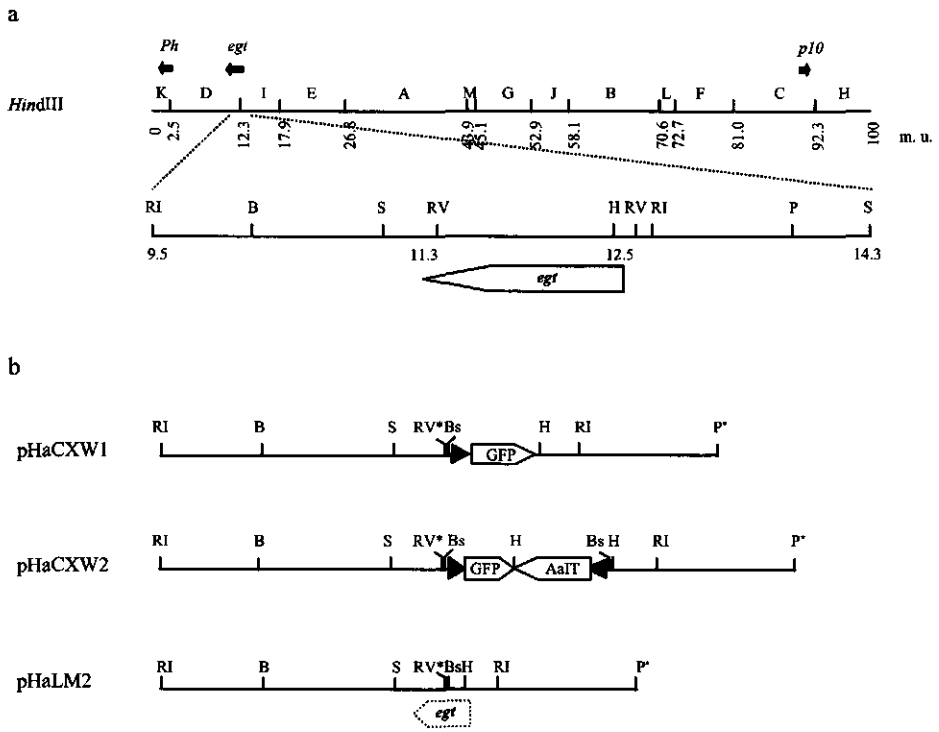


Figure 8-1. (a) Location of *egt* on the HaSNPV genome and a detailed map of the *egt* region (Chen *et al.* 1997a, 2000). (b) Diagram showing the plasmid transfer vectors used to construct recombinant HaSNPVs. The promoter driving the expression of GFP (►) is derived from AcMNPV polyhedrin and that of AaIT (◄) from HearNPV polyhedrin. The restriction sites indicated here are B, *Bam*HI, Bs, *Bsu*36I, RI, *Eco*RI, RV, *Eco*RV, H, *Hin*dIII, P, *Pst*I, S, *Sst*I. * Sites disappearing after blunt ligation.

Isolation of the recombinant viruses

The recombinant HaCXW1, lacking *egt* but containing GFP, was isolated after cotransfection of Hz-AM1 cells with wt HaSNPV DNA and the plasmid pHaCXW1 DNA followed by three rounds of plaque purification. After the initial cotransfection about 2% of the plaques contained GFP, as viewed with an UV microscope. A similar strategy using the plasmid pHaCXW2 was used to generate the recombinant virus HaCXW2 that lacks *egt*, but contains GFP and AaIT. This recombinant contains a *Bsu*36I site at either end of the insert (Fig. 8-1). By cotransfection of Hz-AM1 cells with HaCXW2 DNA, linearized with *Bsu*36I, and the plasmid pHaLM2, the recombinant HaLM2 was generated which lacks an active *egt*, as well as the GFP-AaIT insert present in the parental virus HaCXW2. Viral DNA of the recombinants HaCXW1, HaCXW2 and HaLM2 was compared to wt HaSNPV, but no differences other than those expected at the *egt* locus were detected (data not shown). The structure of the *egt* locus of wt and recombinant HaSNPV was checked by restriction enzyme analysis, PCR and sequence analysis (data not shown).

EGT activity

To determine if HaSNPV and the derived recombinants possess or lack an *egt* gene, EGT activity in supernatants of uninfected or infected Hz-AM1 cells was analyzed by an enzymatic assay using [^{3}H]-ecdysone and UDP-glucose as substrates (O'Reilly *et al.*, 1992). If present, EGT will convert ecdysone (E) to an ecdysone-sugar conjugate (EC). This was observed for wt AcMNPV-infected Sf21 cells (Fig. 8-2, lane 1) as expected and for Hz-AM1 cells infected with wt HaSNPV (Fig. 8-2, lane 4). The supernatants of cells infected with the recombinant viruses HaCXW1 (Fig. 8-2, lane 5), HaCXW2 (lane 6) and AcMNPV-RM2 (lane2) were unable to produce an ecdysone-sugar complex. This indicates that HaSNPV possesses a functional *egt* gene, and that in the recombinant HaSNPVs the *egt* gene has been effectively inactivated. AcMNPV-RM2 lacks the egt gene (Bianchi *et al.*, 2000). Since HaLM2 was derived from HaCXW2, the former recombinant was not included in the assay, as it was not expected to contain an active EGT.

Biological activity of the recombinant viruses

The biological activity of the wt and recombinant HaSNPVs was determined in a droplet-feeding bioassay using second instar *H. armigera* larvae. LC_{50} values of HaCXW2 and HaLM2 were not significantly different from wild type HaSNPV, whereas HaCXW1 had a significantly different LC_{50} value from the wild type (Table 8-1). ST_{50} values of the HaSNPV recombinants were significantly lower than wild type HaSNPV (Table 8-2). In second instar *H. armigera* larvae the *egt* deletion mutants HaCXW1 and HaLM2 exhibited ST_{50} values 26.3% and 28.1% lower, respectively, than wt HaSNPV. The two

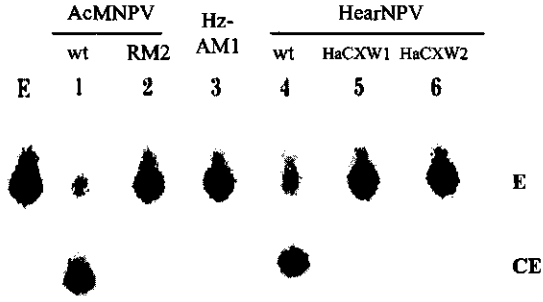


Figure 8-2. EGT activity in supernatants of infected Hz-AM1 and Sf21 cells. Hz-AM1 cells were mock infected (lane 3) or infected with wt HaSNPV (lane 4) or recombinants HaCXW1 (lane 5), HaCXW2 (lane 6). Sf21 cells were infected with AcMNPV (lane 1) or AcMNPV with *egt*-RM2 (lane 2). The position of [^{3}H] ecdysone substrate is marked (E) on the right as are the positions of the conjugated ecdysone (CE) generated by viral EGT activity.

Table 8-1. Dose-mortality relationship of second instar *H. armigera* larvae Infected *per os* with wt HaSNPV and various recombinants

Virus	Log $LC_{50} \pm SD^a$	LC_{50}	Slope $\pm$ SD[a]	χ^2/df
HearNPVwt	4.03 ± 0.04^a	10616	0.85 ± 0.01^a	0.99/3, 2.57/3
HaCXW1	3.49 ± 0.02^b	3092	0.78 ± 0.17^a	1.28/3, 0.31/3
HaLM2	3.95 ± 0.04^a	8912	0.64 ± 0.32^a	0.45/3, 3.23/3
HaCXW2	3.87 ± 0.28^a	7413	0.80 ± 0.44^a	2.11/3, 0.27/3

[a] Different superscript letters indicate significant differences between treatments (P = 0.05).

HaSNPV *egt* deletion mutants have a similar killing speed of approximate 69 h. The largest average reduction in the time required for the virus to kill *H. armigera* was found with recombinant HaCXW2, which contained an AaIT gene in the *egt* locus. The ST_{50} value of HaCXW2 was 63.9 h, which is a reduction of about 32% compared to wt HaSNPV. However, the killing speed of HaCXW2 was not significantly different from HaCXW1 and HaLM2 (Table 8-2). The pathologies of HaCXW1 and HaLM2 infections, on the one hand and HaCXW2 infection on the other, were quite different. Infection with the latter showed the typical signs of paralysis as reported for AcMNPV recombinants expressing AaIT (Stewart *et al.*, 1991; Maeda *et al.* 1991).

DISCUSSION

The *egt* gene was first identified in AcMNPV (O'Reilly and Miller, 1989) and the EGT enzyme is believed to inactivate host ecdysteroids by conjugation of a sugar, either glucose or galactose, to the ecdysteroids. As a result molting and pupation of the infected host are blocked (O'Reilly, 1995). *Egt* homologues have been identified in a good number of baculoviruses including HaSNPV (Chen *et al.*, 1997b). EGT activity analysis in the supernatant of HzAM1-infected insect cells indicated that HaSNPV encodes a functional EGT that can transfer sugar to ecdysteroids and inactivate the ecdysteroids (Fig. 8-2). This observation paved the way to generate an *egt* deletion mutant of HaSNPV to improve its insecticidal properties.

Deletion of the *egt* gene from AcMNPV resulted in a recombinant that killed its target hosts, like *S. frugiperda*, *Trichoplusia ni* and *H. virescens*, faster than the wt AcMNPV. (O'Reilly and Miller, 1991; Treacy *et al.*, 1997). However, a similar *egt* deletion mutant did not kill *S. exigua* larvae more rapidly in bioassays (Bianchi *et al.*, 2000) suggesting that the responses in insect hosts are not always predictable and need evaluation. An *egt* minus *Lymantria dispar* MNPV (LdMNPV) strain exhibited a reduced ST_{50} value of about 33% as compared to wt-LdMNPV, but only in fifth instar larvae (Slavicek *et al.*,

1999). A similar reduction (26 - 28%) in the speed of kill was observed for the HaSNPV *egt* deletion mutants (Table 8-2) in second instar *H. armigera* larvae. An *egt* deletion mutant of *H. zea* SNPV (HzSNPV), which is closely related to HaSNPV according to *egt* and *polyhedrin* gene sequence data (Chen *et al.,* 1997a, b), however, did not enhance the killing speed of the virus in *H. zea* neonate larvae. This may be due to the quick action of HzSNPV in this instar (Popham *et al.,* 1997). Apparently, the efficacy of *egt* minus baculoviruses is dependent not only on the virus replication rate, but also on the host and host developmental stage.

The AaIT-expressing recombinant HaCXW2 had a reduced ST_{50} relative to wt HaSNPV, but was only slightly more effective than the *egt* deletion mutant HaCXW1 (Table 8-2). This may be due to the fact that the polyhedrin promoter was used, which is expressed very late after infection. Promoter-dependent effects on the expression of insect-specific neurotoxin genes and larval paralysis have been previously reported for AcMNPV

Table 8-2. Time-mortality relationship of second instar *H. armigera* larvae infected *per os* with wt HaSNPV and various recombinants

Virus	$ST_{50} \pm SD^a$	Slope $\pm$ SD[a]
HaSNPVwt	94.4 ± 1.53^a	5.31 ± 0.49^a
HaCXW1	69.5 ± 2.12^b	7.24 ± 2.93^a
HaLM2	67.9 ± 5.16^b	6.73 ± 2.11^a
HaCXW2	63.9 ± 3.54^b	5.99 ± 0.04^a

[a]Different superscript letters indicate significant differences between treatments (P = 0.05).

(Tomalski and Miller, 1991; Lu *et al.*, 1996). Application of both early and late promoters driving the expression of toxins could thus further improve the efficacy of HaSNPV. The viral basic protein (*p6.9*) and the *Drosophila* hsp70 promoter were considerably more effective than the *polyhedrin* promoter in reducing the speed of action of AcMNPV against both *T. ni* and *S. frugiperda* (Lu *et al.*, 1996). The ST_{50} value does not entirely indicate the real effectiveness of a given recombinant virus especially for those viruses expressing insect neurotoxins. Recent experiments with the HaSNPV recombinants, described here, also showed that the food consumption of insects infected with the AaIT-expressing HaSNPV (HaCXW2) was much less than those infected with wt HaSNPV or the *egt* deletion mutants (HaCXW1) (Sun *et al.*, 2001).

The green fluorescent protein (GFP) has been used successfully as an effective reporter gene in many heterologous expression systems. GFP has been expressed along with

baculovirus genes and has been used as a marker to select recombinants (Chao *et al.*, 1996; Wilson *et al.*, 1997). Here GFP also proved to be a helpful marker in the screening of HaSNPV recombinants. However, when larvae were infected with HaCXW1 and HaCXW2 the fluorescence was minimal upon illumination with UV light. A similar observation was made with larvae of the diamondback moth, *Plutella xylostella* (Chao *et al.*, 1996) and *T. ni* (Barrett *et al.*; 1998) when infected with recombinant GFP-expressing AcMNPV. This may be due to the relatively dark skins of these larvae quenching the signal.

Expression of GFP did not change the biological activity of the HaSNPVs, as HaCXW1 and HaLM2, which differ only in the presence or absence of GFP, did not show significantly different LC_{50} and ST_{50} values in the bioassays. A similar result has been reported for LdMNPV (Bischoff and Slavicek, 1999), where the polyhedral envelope protein (*pep*) gene was deleted and replaced by GFP. However, in the latter case GFP supposedly prevented virus-induced larval liquefaction. This was not observed for HaSNPV recombinants infecting *H. armigera* larvae.

The availability of GFP-containing HaSNPV recombinants now allows the analysis of pathological effects of these viruses in target and non-target hosts, competition experiments between the wt and recombinant viruses and experiments monitoring the fate of these recombinants in the field. Furthermore, field experiments carried out in 1999 on a cotton field in China already indicated that both HaCXW1 and HaCXW2 recombinants killed *H. armigera* faster than the wt HaSNPV (Sun *et al.*, 2001). An expanded field trial will be conducted in 2000 to determine the effectiveness of these genetically modified HaSNPVs on a large scale in cotton and their effect on cotton yield.

ACKNOWLEDGEMENTS

We thank Magda Usmany and Els Roode for technical assistance and Felix Bianchi for assistance in the statistical analysis of the bioassay data. This research was supported by a joint grant from the Chinese Academy of Sciences and Royal Academy of Sciences of The Netherlands (contract 97CDP010), a grant from the Ministry of Science and Technology of China, grants from the National Natural Science Foundation of China and from the Dutch-Israeli Agricultural Research Programme (DIARP) 97/29. X. Chen is a recipient of a PhD Sandwich Fellowship from Wageningen University.

Chapter 9

GENERAL DISCUSSION: BACULOVIRUS GENOME EVOLUTION

INTRODUCTION

Baculoviruses are widely used for insect pest control and as vectors for high level expression of interesting polypeptides in insect cells. The recent development of recombinant baculoviruses with improved insecticidal properties provides the potential to successfully compete with chemical pesticides in the commercial market. Baculoviruses also hold promise as gene delivery vectors for gene therapy. All these applications call for detailed insight and understanding of the molecular and genetic basis of the viral infection process, such as entry, gene expression, genome replication and virion assembly. With the ultimate aim to improve HaSNPV as a viral insecticide, the research presented in this thesis first describes the genetic organization and sequence of the HaSNPV genome. This information may explain the unique features of this virus and add to the understanding of the biological properties of this virus. Also, recombinant HaSNPVs with a faster speed of action were finally constructed and their properties and potential use for bollworm control in cotton are described in this thesis.

Genetic relatedness of baculoviruses often has been based on phylogenetic analysis of a single selected gene, such as the occlusion body protein (polyhedrin) and the DNA polymerase (Zanotto *et al.*, 1993; Bulach *et al.*, 1999; Hu, 1998; Bideshi *et al.*, 2000; Moser *et al.*, 2001). Those analyses indicated that baculoviruses cluster into two groups, NPVs and GVs, while NPVs were further subdivided into Groups I and II. The branch position of the GVs and Group II NPVs has not been clearly resolved. Also the phylogenetic and taxonomic implications of the S and M phenotype remained unresolved mostly due to the lack of detailed studies on SNPVs and sequence data. The sequence of the HaSNPV genome tries to fill this gap in our knowledge. With the complete genome sequence of some pro- and eukaryotic organisms, one could use this information to make a predication on a gene product based on sequence, including predicting protein structure and possible protein-protein interactions (Huynen *et al.*, 2000). Different approaches have also been pioneered to analyze the phylogenies based on complete genome sequence data (Kondrashov, 1999, for review). The availability of complete genome sequence data for a growing number of baculoviruses, including those of HaSNPV and HzSNPV as presented in this thesis, has led to an interest in the use of such

genomic data for phylogenetic analyses. In this chapter, comparative genomics and phylogeny of baculovirus based on their genes or genomes and some future perspectives on the use of recombinant baculoviruses as bio-insecticide are discussed.

SNPV GENOMES ARE SIMILAR TO MNPV GENOMES

Among the baculoviruses, the complete sequences of 5 MNPVs and 3GVs have so far been determined (Table 9-1). The genome sequence data of SNPVs (HaSNPV and HzSNPV) presented in this thesis now allow a meaningful comparison of the genome organization of the two morphotypes of baculoviruses. This also provides a wealth of information to determine the distinctive features, the diversity as well as the evolution of baculovirus genomes.

The complete sequences of the HaSNPV and HzSNPV genomes indicated sizes of 131,403 bp and 130,678 bp, respectively, potentially encoding 135 and 139 ORFs. Five *hr*s were found in both genomes (Chapters 6 and 7). The genomes of these SNPVs have, in general, quite a similar gene content and genome structure as MNPVs and GVs. The ORFs are tightly packed with minimal intergenic regions and their orientations are almost evenly distributed along the genome. Both strands of the genome are involved in coding functions and generally the genes are most likely to be unspliced. Gene classes (early, late, very late) do not appear to be specifically clustered in baculovirus genomes. Gene duplications are also observed in HaSNPV/HzSNPV genomes. The diversity of baculoviral genomes, including those of HaSNPV and HzSNP, is exhibited in the nucleotide (G+C) composition, the size, and in the number of unique ORFs, repeated ORFs and *hr*s (Table 9–1). Direct genome comparison of HaSNPV and HzSNPV to MNPVs and GVs using GeneParityPlot (Hu *et al.*, 1998) indicated that

Table 9-1. Characteristics of different baculovirus genomes

Characteristic	AcMNPV	BmNPV	OpMNPV	LdMNPV	SeMNPV	HaSNPV	HzSNPV	CpGv	PxGV	XcGV
Size(kb)	133.9	128.4	132	161	135.6	131.4	130.7	123.5	101	178.7
Coding (%)	90	90	89	87	90	87	87	NK	86	88
Total ORFs	154	136	152	163	139	135	139	143	120	181
Unique ORFs	14	4	26	47	17	20(0)*	24(4)*	26	16	52
Number of *hr*	9	7	5	13	4	5	5	NK	4	9
G+C content	41	40	55	58	44	39	39	NK	40	41
Classification	NPV I	NPV I	NPV I	NPV II	NPV II	NPV II	NPV II	GV	GV	GV

AcMNPV, Ayres *et al.*, 1994; BmNPV, Gomi *et al.* 1999; OpMNPV, Ahrens *et al.*, 1997; LdMNPV, Kuzio *et al.*, 1999; SeMNPV, IJkel *et al.*, 1999; HaSNPV, Chen *et al.* 2001; PxGV, Hashimoto *et al.*, 2000; XcGV, Hayakawa *et al.*, 1999; HzSNPV, this thesis; CpGV, O'Reilly, personal communication. NK, not known.

* 20 unique HaSNPV ORFs have homologs in the HzSNPV genome

HaSNPV/HzSNPV have a distinct gene order but are closely related to group II NPVs, SeMNPV and LdMNPV. This also suggests that the MNPV and SNPV morphotype can not be regarded as a phylogenetic marker.

HOMOLOGOUS REGIONS: THE MOST HYPER-VARIANT REGIONS

Five homologous regions (*hr*s), the first *hr*s of a SNPV, were found in the HaSNPV genome (Chapter 6). This clearly shows that the presence of *hr*s is a unique feature of all baculovirus genomes (Table 9-1). *Hr*s occur at multiple locations along the genome, ranging from 4 in PxGV to 13 in LdMNPV. It has been demonstrated that *hr*s serve as putative origins of DNA replication in transient replication assays (Kool *et al.*, 1995) and as enhancers of RNA polymerase II-mediated transcription (Guarino & Summers, 1986; Guarino *et al.*, 1986). It has also been reported that the conserved imperfect palindromes within the repeated sequences are critical for their functions (AcMNPV). However, no long palindromic sequence was found in the HaSNPV/HzSNPV *hr*s. Instead, these *hr*s are characterized by two types of direct repeats, similar to the XcGV *hr*s, which also contain only imperfect direct repeats and have no palindromic structure (Hayakawa *et al.*, 1999). In contrast to most of the ORFs, *hr*s do not share significant sequence identity among different baculoviruses. Hence, *hr*s diverge significantly among baculoviruses both in sequence and in structure.

Comparison of the HaSNPV and HzSNPV sequences strongly supports the view that they are two variants of the same virus species. The major insertion and deletion as well as the lowest sequence identity between these two genomes are found in or near the *hr* regions. Also, most of the unique HzSNPV ORFs map in the flanking regions of *hr*s. This supports the view that in different variants of the same viral species *hr*s are hypervariable (Muñoz *et al.*, 1998). This observation suggests that *hr*s might also be hot spots of recombination within or between baculovirus genomes. In support of this view, the most distinct regions among different baculovirus genomes are flanked by or contain a disproportionate number of *hr*s (Hayakawa *et al.*, 2000).

CONSERVATION OF BACULOVIRUS GENE CONTENT

Based on ten baculoviral genome sequences, including those of HaSNPV and HzSNPV, 63 ORFs were found conserved among all baculoviruses, suggesting that they are required for the basic baculovirus features (Table 9-2). So far 36 of these 63 core genes have been assigned putative functions, while the functions of the remaining 27 ORFs are still unknown. Seventeen of the former 36 genes encode proteins for DNA replication, regulation of transcription and late gene expression, *e.g.* genes essential for DNA replication such as DNA polymerase, helicase, *ie-1*, *lef*-1, *lef*-2 and *lef*-3 (Lu *et al.*, 1997). A second set of genes (15 of the 63) encodes structural proteins, including components of polyhedra, capsid, tegument and envelope. Four genes that appear to provide the virus

with a selective advantage in nature function at the cellular or organism levels. It is worth

Table 9-2: Function of the 63 conserved baculovirus genes.

Replication / expression	Structure	Auxiliary	Unknown
39K, dbp1, dnapol,	*ac23, fp25K, gp41, odv-e18,*	*alk-exo, fgf, sod, ubiquitin*	*38K, 38.7K, ac22, ac29, ac38, ac53, ac66,*
helicase, lef1, lef2 , lef3,	*odv-e25, odv-ec27, odv-e56,*		*ac68, ac75, ac76, ac78, ac81, ac82, ac92,*
lef4, lef5, lef6, lef8, lef9,	*odv-e66, p6.9, p74, p95,*		*ac93, ac96, ac106, ac109, ac110, ac115,*
lef11, ie1, me53, p47, vlf1	*pk1, polh, vp39, vp1034*		*ac119, ac142, ac145, ac146, p12, p40, p45*

noting that the genes involved in DNA or nucleotide metabolism, *e.g.* DNA ligase, ribonucleotide reductase (two subunits) and dUTPase, are not conserved in baculoviruses. In herpesviruses, which have similar genome size (100 kb-240 kb), only about 40 genes are conserved (Albrecht *et al.*, 1992). We therefore may expect that for baculoviruses the minimal number of conserved genes may be less than 63, which then represent the minimal requirement for sustaining the baculovirus life cycle. Hence the 'minimal' baculovirus genome could be around 40-50 kb in size.

So far 12 ORFs were found to be unique for NPVs as they were absent from GVs. These ORFs might correspond to unique morphological or pathological properties of NPVs. Two viral structural genes, *p80/p87*and *pp34,* are only found in NPVs, P80/P87 being associated with both BV and ODV (Lu and Carstens, 1992; Müller *et al.,* 1990). PP34, an integral component of the polyhedron envelope/calyx structure of NPVs (Gombart *et al.*, 1989), is absent from GVs which underscores the difference in the structure of GV and NPV occlusion bodies (Federici, 1986). A homolog of *arif*-1 (Dreschers *et al.*, 2001; Roncarati and Knebel-Mörsdorf, 1997), which is involved in the rearrangement of the actin cytoskeleton of the infected cell, is only found in NPVs. It may contribute to the difference in pathology between GVs and NPVs (Federici, 1986).

GENE ACQUISITION: REPEATED GENES

The baculovirus genome size may vary from 100 kb (PxGV) to 180 kb (XcGV) (Table 9-1), even though the gene contents are more or less the same. Although processes, such as deletion and insertion, transposition and horizontal transfer of genes, might be involved, the major contributor to the genome size difference is gene duplication. The largest genome, that of XcGV, contains 30 repeated genes accounting for 37.5 kb or 20% of the genome. Similarly LdMNPV contains 32 repeated genes that account for 27.5 kb or 17% of the genome. HaSNPV, which has an average genome size, contains only 5 gene

repeats, two *iap* and 3 *bro* genes (Hayakawa *et al.*, 2000; Table 9-3).

So far 17 genes have been found to be repeated in different baculoviruses (Table 9-3). Three of these, *iap*, *odv-e66* and *fgf*, are conserved in all baculoviral genomes so far. The first member of the IAP family, consisting of proteins which act as "inhibitors of apoptosis", were found in granulovirus (Crook *et al.*, 1993). It is notable that most baculoviruses have more than one copy of *iap*, suggesting that these genes are functionally important. They might have been picked up from different genetic sources and may be active in different apoptotic pathways or in specific cell types, tissues, or species. HaSNPV induces apoptosis in different insect cells, like Tn-Hi5, Sf21 and Se-UCR, (Dai *et al.*, 1999; Chen *et al.*, unpublished data). Future experiments will reveal if both *iaps* of HaSNPV are functional in a differential manner. The most highly repeated gene family, *bro*, is present from one (AcMNPV) to 17 copies (LdMNPV), but is absent in SeMNPV and CpGV. Functional analysis indicated that different *bro* genes might have different functions (Kang *et al.*, 1999; Zemskov *et al.*, 2000). Phylogenetic analysis also showed that *bros* may have been acquired by either duplication or from a different source, and have been adapted to their respective functions during the viral evolution (chapter 7).

Over all, gene duplication appears to have been a major driver in the generation of the largest baculoviral genomes. There is no observable relationship between genome size and baculovirus phylogeny. It has been suggested that in bacteria, increases in genome size have occurred independently during the evolution of different lineages (Graur and Li,

Table 9-3. Baculovirus repeat genes

Gene	AcMNPV	BmNPV	OpMNPV	LdMNPV	SeMNPV	HaSNPV	HzSNPV	XcGV	PxGV	CpGV
iap	2	2	4	2	2	2	2			3
bro		5	3	16	-	3	3	8	-	
Ac145/150	2	2	2	2	3			4	3	
ctl			2	2	-	-	-		-	-
ptp			2	-	-	-	-	-	-	2
dbp				2						
enhancin	-	-	-	2	-	-	-	4	-	-
rnr	-	-		2		-	-			2
ld151/ld162	-	-	-	2	-	-	-	-	-	-
ld34/163	-	-	-	2	-	-	-	-	-	-
odv-e66					2					
p26					2			-	-	-
p10								3	3	-
fgf								2	2	
cpgv16l	-	-	-	-	-	-	-	2	2	2
xcrep1	-	-	-	-	-	-	-	5	-	-
xcrep2	-	-	-	-	-	-	-	2	-	-
total	4	9	13	32	9	5	5	30	10	9

- : Gene absent in the genome

UNIQUE ORFs

So far 20 potential ORFs are unique to HaSNPV/HzSNPV and 4 additional ORFs unique to HzSNPV. Presumably, these ORFs may contribute to unique features such as host range and virulence. With 10 completely sequenced baculoviral genomes, so far 220 ORFs are found only in one baculoviral species (Table 9-1). Only very few of these have been analyzed to some degree. For example the SeMNPV unique gene Se117 encodes an ODV-specific nucleocapsid protein (IJkel *et al.,* 2001). Three other unique SeMNPV genes (Se116, Se17 and Se18) are known to be expressed in virus infected cells, but their functions are still enigmatic (IJkel *et al,* 2001; IJkel, 2001). Both Op25 and Op32 of OpMNPV are expressed early in infection, but deletion of either gene from the genome did not affect OpMNPV replication *in vitro* (Shippman *et al.,* 1997).

Analysis of the HaSNPV/HzSNPV genome sequence indicated that most unique genes and ORFs have significantly different codon usage and a higher synonymous and non-synonymous substitution rate in comparison to the conserved genes (Chen *et al.,* unpublished). This may suggest that the putative genes have been acquired later during baculovirus evolution. Further experimental research would reveal if those unique genes were functional or not.

POSSIBLE GENETIC DETERMINANTS OF THE SNPV AND MNPV PHENOTYPES

Little has been known about the molecular genetic mechanism governing the SNPV versus MNPV morphotype. Swapping of the *polyhedrin* and *p10* genes between SNPV and MNPV, or GV and MNPV, indicated that these genes do not play a role in these characteristics (Van Oers *et al.,* 1998; Eason *et al.,* 1998; Hu *et al.,* 1999). It is conceivable that one or more of the unique genes, alone or in concert with conserved genes, are responsible for this morphogenetic effect. The HaSNPV genome sequence did not clearly show any HaSNPV gene being directly involved in this phenomenon. Genomic comparison of HaSNPV/HzSNPV with MNPVs and GVs indicated that Ac120 homologs (Bm98, Op120, Ld155a and Se37) are present in all MNPVs, but are absent in SNPVs and GVs. Ac120 potentially encodes an 82 amino acid-long protein with a high amount of hydrophobic amino acids and a high isoelectric point (pI = 10). Since Ac120, and its homologs, are unique to MNPV so far, it may be a candidate gene involved in the determination of the phenotypes either in SNPVs or GVs. Also, future experiments will reveal, if any such HaSNPV unique ORF has such a function.

FUNCTIONAL PREDICTION AND ANALYSIS OF BACULOVIRUS GENOME

The HaSNPV genome contains 135 ORFs, of which only about 60 ORFs have homologs with assigned functions in other baculoviruses. In other baculoviruses, the function of only a limited numbers of genes has been analyzed. The sequencing of complete genomes has created the possibility of not only to analyze the role of the gene product but also to

predict the protein function from the genomic context. Based on the co-localization of genes on the genome, different approaches have been used to predict gene function by using either the fusion of genes, the local neighbourhood or merely the presence of genes in the same genome (Marcotte *et al.,* 1999; Enright *et al.,* 1999; Dandekar *et al.*, 1998; Huynen and Bork, 1998; Pellegrini *et al.*, 1999). In contrast to classical homology-based function prediction, methods based on genomic context do not directly predict the function of the gene product, but rather the existence of functional interactions between gene products.

Even though many conserved gene clusters have been determined in baculovirus genomes by using GeneParityPlot analysis (Hu *et al.*, 1998; IJkel *et al.*, 1999; Hayakawa *et al.*, 2000; chapter 6), functionally related genes are not always clustered together. For instance, the genes involved in viral DNA replication, including DNA polymerase, helicase, *ie-1*, *lef-1*, *lef-2* and *lef-3,* are "randomly" distributed along the genome. Interaction within a conserved gene set might also occur in different ways or at other levels, *e.g.* on mRNA, if not at the protein level (not enough proof yet). This hypothesis is supported by the transcription mapping analysis of a 11.3 kb SeMNPV DNA fragment, which contains a highly conserved baculovirus region (Heldens *et al.,* 1998). Further evidence is furnished from transcriptional analysis of the AcMNPV gene clusters having 3' co-terminal transcripts (Friesen and Miller, 1985; Lübbert and Doerfler, 1984). Detailed analysis should reveal functional interactions among gene products of the conserved gene cluster in baculovirus genomes.

DNA arrays provide a novel means to measure the expression of hundreds or thousands of genes simultaneously as well as allowing high-throughput characterization of samples (Lockhart and Winzeler, 2000). Baculoviruses are ideal candidates for DNA array technology, as every viral gene can be printed onto a single array. Therefore, the complete gene expression profile of an entire genome can be elucidated from the results of a single array (Mohr and Rihn, 2001; Jenner *et al.*, 2001; Zhu *et al.,* 1998). A HaSNPV DNA array can now be designed to analyze the complete gene transcription pattern of the HaSNPV genome *in vitro* as well as *in vivo.*

BACULOVIRUS PHYLOGENY BASED ON SINGLE GENES

For many years baculovirus phylogeny relied mostly on OB protein sequence. Those early studies suggested that lepidopteran NPVs evolved from a common lepidopteran baculovirus ancestor rather than from a cross-infecting virus from different orders of arthropods (Rohrmann *et al.*, 1981; Vlak and Rohrmann, 1985). It has been postulated that GVs have evolved from lepidopteran NPVs in the early evolutionary stage prior to the divergence of the two major clades of lepidopteran NPV (namely, Group I and Group II) (Zanotto *et al.*, 1993; Cowan *et al.*, 1994). Preliminary analysis using an evolutionary

clock suggested that lepidopteran baculoviruses may not have co-evolved with their insect host species (Hu, 1998).

However, the OB protein gene is not the ideal gene for phylogenetic analysis because of its small size and relative high sequence conservation thus providing limited phylogenetic information. Moreover, since baculoviruses contain more than 100 ORFs, phylogeny based on a single gene may not accurately represent baculovirus relatedness. It is therefore of interest to check if different genes would produce similar or different trees. Besides the structural proteins such as polyhedrin, one baculovirus auxiliary gene, *egt,* and one essential gene, *lef*-2, involved in viral DNA replication and gene expression, were tested and compared (Chapter 2 and Chapter 3, Fig. 9-1). Even though these genes have quite different functions and might have experienced different selective pressure during viral evolutionary history, their phylogenetic trees have a similar topology. This similarity in trees, using different genes, is further substantiated by the results obtained with the DNA polymerase gene (Bulach *et al.,* 1999). All these analyses support the view that baculoviruses fall into two separated clades, GV and NPV, whereas the NPV clade further sub-divides into Group I and II. With those analyses some further splits within subgroups in Group I and Group II NPV have been hypothesized (Chapter 4; Hu *et al*, 1998; Bulach

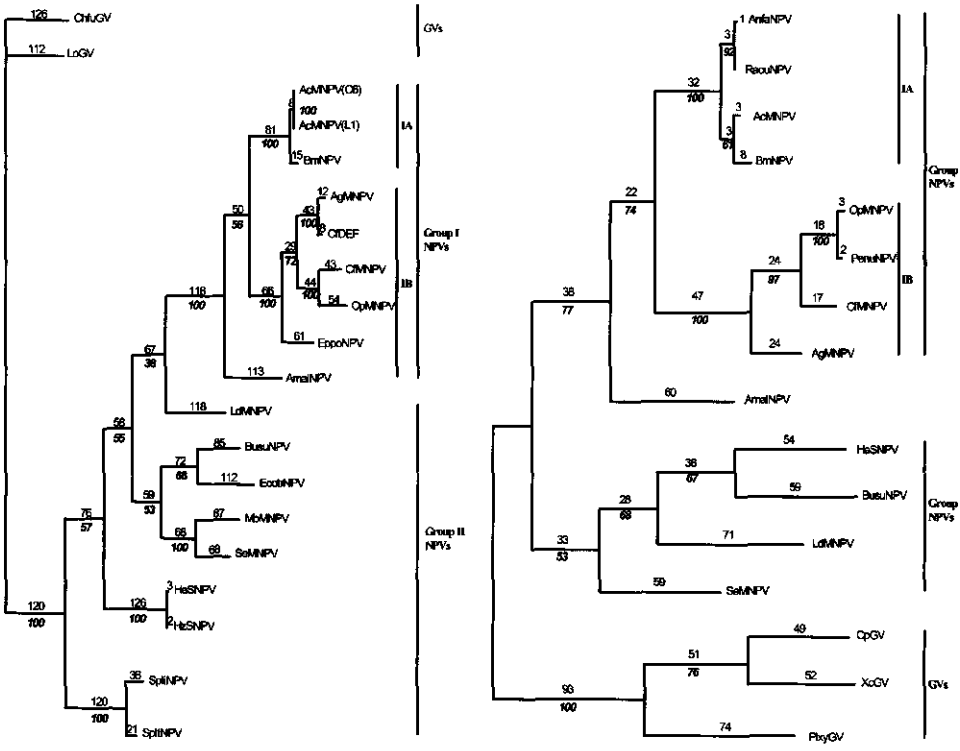


Figure 9-1. Phylogenetic trees of baculovirus based on their EGT and LEF-2 amino acid sequence.

et al., 1999). However deeper branches of the group II NPV clade are not yet resolved with high fidelity.

BACULOVIRUS PHYLOGENY BASED ON GENE COMBINATION

Compared with single genes, combining genes may reduce sampling errors in phylogenetic analysis and causes the phylogenies to converge towards the correct solution with good support (Mitter *et al.*, 2000). Combined alignment of 63 conserved genes (Table 9-2) from ten baculovirus genomes (Table 9-1), was used to reconstruct a phylogenetic tree (Herniou *et al.*, 2001). A single most parsimonious tree was constructed and all branches were supported with high bootstrap value (Fig 9-2a). This tree is in accordance with most of the single gene trees (Herniou *et al.*, 2001) suggesting that baculoviral genomes are monophylectic. This analysis also supports the view that obtaining characters from independent genes would hold greater benefit for phylogeny inference than obtaining the same total number of characters from a single gene (Cummings *et al.*, 1995; Graybeal, 1998).

BACULOVIRUS PHYLOGENY BASED ON GENOMIC ORGANIZATION

GeneParityPlot has been developed to compare and analyze baculovirus genomic organization (Hu *et al.*, 1998). With this approach different conserved baculovirus gene clusters have been detected (IJkel *et al.*, 1999; Hayakawa *et al.*, 2000; Chapter 6). In addition, direct comparison of whole genomes with this method primarily indicated that genome organization, including the gene content and gene order, could be an independent parameter character in addition to single gene or protein sequences, to be used directly for phylogenetic analysis. However, this analysis is only qualitative showing gene clusters (Hu *et al.*, 1998). In the following we try to construct baculovirus phylogenetic trees (genomic trees) based on their genome organization, both on gene order and gene content.

The approach, which is applied here, infers phylogeny by using gene order and is termed Neighborhood Disruption Frequency Analysis (NDFA) and this is compared to the breakpoint distance analysis outlined by Blanchette *et al.* (1999). This NDFA analysis holds that for each pair of genomes for all genes shared by these genomes the pairs of genes in one genome that were disrupted in the other were counted. This results in a data matrix after pair-wise comparison alignments of all nine genomes (data not shown). Similar to breakpoint analysis, NDFA is independent of the mechanism of gene rearrangement. The genomic distance matrices then are analyzed by maximum parsimony. A single tree was produced, presented in Figure 9-2b. The topological structure of this genomic tree is similar to that of the combined tree of the 63 conserved genes and the gene order trees, which Herniou *et al* (2001) recently constructed by using relative breakpoint

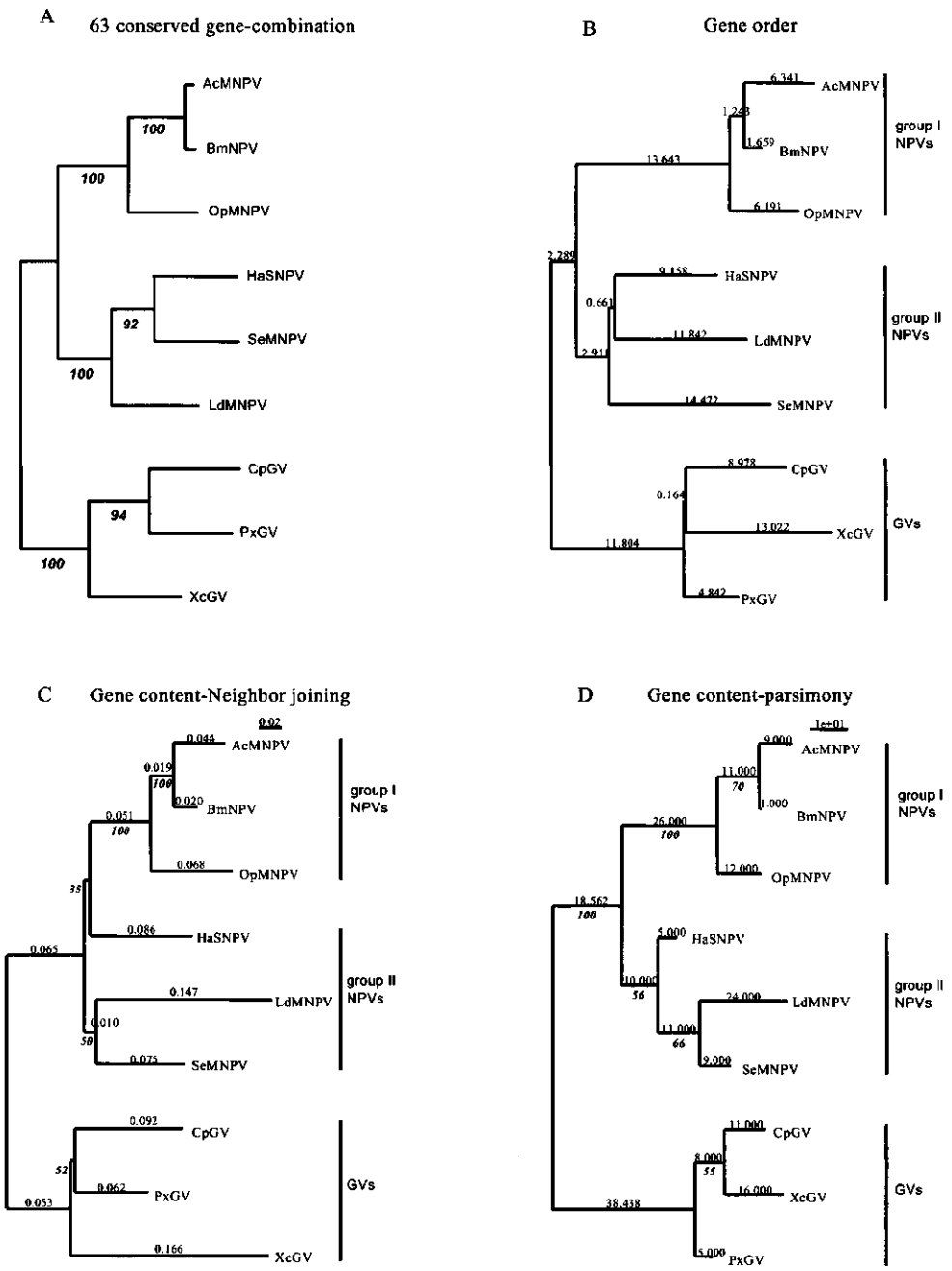


Figure 9-2 Baculovirus phylogenetic trees based on nine completely sequenced baculovirus genomes. A. Most parsimonious tree of the combined sequence of the 63 conserved genes (from Herniou *et al.*, 2001). B. Baculovirus phylogenetic tree based on shared gene order. Neighbor-joining (C) and most parsimonious (D) tree based on baculovirus gene content. The number above the line indicates the distance. The bootstrap value (100 replicates) is shown in italics.

distance and neighbor pair analysis (but different from the relationships among the group II NPVs and /or GVs.

BACULOVIRUS PHYLOGENY BASED ON GENE CONTENT

Gene content trees have been constructed for cellular genomes by several groups using different methods (Snel *et al.*, 1999; Fitz-Gibbon and House, 1999; Tekaia *et al.*, 1999). Recently, Montague and Hutchison (2000) reconstructed the complex evolution history of herpesviruses based on the genome content by using COG (clusters of orthologous groups) data. For this method, COGs were first identified among the different genomes (Tatusov *et al.*, 2001). Then a matrix was generated recording the presence or absence of each COG as 1 or 0 in each genome. Phylogenetic analyses were then performed by using the normal phylogenetic analysis program such as PAUP.

To construct a baculovirus genomic tree based on genomic content, orthologous gene families were constructed (Tatusov *et al.*, 2001). Repeated genes, such as *bro* genes, are not included in this analysis. Then a multiple genome alignment was generated recording the presence or absence as 1 or 0 of each baculovirus gene. In this matrix, the conserved genes across all genomes as well as the unique genes to an individual genome are phylogenetically not informative and hence not included. All differentiation is based only on the partially conserved genes. By running PAUP 4.0 two trees either neighbor joining or parsimony were constructed based on the alignment data. Branch support was assessed by bootstrap analysis (fig. 9-2c/d).

Distance and maximum parsimony analysis of the genome content results in two trees (Fig. 9-2c and 9-2d) which are comparable to that predicted by phylogenetic analysis of the combined 63 conserved genes (Herniou *et al.*, 2001) and gene order of the genomes (Fig. 9-2b). However, the relative branching order of Group II NPV and GV remains unresolved. In the distance tree, HaSNPV even separated from LdMNPV and SeMNPV. This branching, however, was not supported by the bootstrap analysis. The ambiguity of the relative position of these species might be due to different reasons such as the long terminal branches effect (Felsenstein, 1978). Both Group II NPV and GV contain a large genome (LdMNPV in Group II NPV, and XcGV in GV), but the major problem might be the lack of sufficient genome data for group II and GV.

HOST RANGE MIGHT NOT BE A CRITICAL CHARACTER FOR BACULOVIRUS TAXONOMY

Most baculoviruses are host specific suggesting that they have been adapted to their respective hosts for optimal survival. Baculoviruses are usually named after the insects from which they were isolated, and the natural host range has been accepted as one of the key critical characters for their classification (Murphy *et al.*, 1995; Blissard *et al.*, 2000).

The molecular basis of host range determination, which may become apparent in different viral replication stages, *e.g.* viral entry, DNA replication and gene expression, however, is not very clear yet. Some genes, like helicase (Maeda *et al.*, 1993, Croizier *et al.*, 1994), *lef*-7 (Lu and Miller, 1995), *hcf*-1 (Lu and Miller, 1995), *hrf*-1 (Thiem *et al.*, 1996), *iaps* and *p35* have been implicated in baculovirus host range (Thiem, 1997, for review). The potential mechanisms to contribute to host range determination may appear to be very diverse.

HzSNPV and HaSNPV (or HearNPV), which were independently isolated from different insect hosts, have been accepted for a long time as two distinct viruses based on their different original insect host (Murphy *et al.*, 1995). However, the genetic data included in this thesis, combined with the biological features, *e.g.* natural host range, morphology, pathology, indicate that they are two variants of the same virus species. This illustrates again that the baculovirus classification and nomenclature system based on the host is often not supported by molecular data. The same situation exists with other presumed baculovirus species, OpMNPV, ChroNPV and PenuNPV; AcMNPV, AnfaNPV, GameNPV, RaouNPV, SpexNPV and TnMNPV. Even though they were isolated from different insect species, many of these viruses might just appear to be different strains of the same virus species, simply because the virus has a broad host range (Blissard *et al.*, 2000). On the other hand, distinct virus species may have been isolated from the same insect species such as HaSNPV and HaMNPV from *H. armigera*, OpMNPV and OpSNPV from *O. pseudotsugata*. The sequence information as presented in this thesis calls for a major revision of baculovirus nomenclature (Van Regenmortel *et al.*, 2000) to better reflect the molecular phylogeny.

IMPACT OF GENOMIC SEQUENCE INFORMATION ON BACULOVIRUS TAXONOMY

The gene sequences and genomic organization of viruses may also be considered as phenotypic traits. It is generally found that the overall degree of phenotypic difference observed between organisms, which also includes sequence divergence, is roughly proportional to the amount of evolutionary distance. Phenetic species defined in terms of overall similarity are often very similar to the species defined as lineages of an ancestral-descendant population. Extensive sequence comparison among the sequences available in the database, either genome sequences or coat protein sequences of potyviruses, for example, have shown the successful application of quantitative taxonomy based on the sequence homologous to all members of the family, resulting in a clear-cut distinction between each of the different taxonomic levels: strains, species, and genera (Shukla *et al.*, 1994; Van Regenmortel *et al.*, 2000).

It has been shown in this thesis that closely related baculoviruses are distinguished from

each other using homology analysis of their genomes and genes. The overall ORFs between nucleopolyhedroviruses and granuloviruses have on average less than 40% amino acid sequence identity while different members of the same genera, either *Nucleopolyhedrovirus* or *Granulovirus*, have a higher identity (>40%) (Chapter 6). Even though HaSNPV and HzSNPV, which were assigned as different isolates of the same virus species, with 98% identity over all ORFs, AcMNPV and BmNPV, which share 93% identity regarding all ORFs, were assigned as distinct species status based on an important biological feature, a different host range. It is proposed here that viruses with genome homology levels of 93% or less should be assigned as distinct virus species. If their homology level is less than 40%, they belong to different baculovirus genera. The latter may apply for *Culex nigripalpus* baculovirus (Moser *et al.*, 2001). More sequence data will probably result in more precise breakpoints.

Table 9-4. List of sequenced polyhedrin/granulin

Virus	Abbreviation	Accession
Autographa californica MNPV (E2)	AcMNPV-E2	K01149
A. californica MNPV (C6)	AcMNPV-C6	L22858
Anagrapha falcifera MNPV	AnfaNPV	U64896
Amsacta albistriga npv	AmblPV	AF118850
Anticarrisa genmatalis MNPV	AgMNPV	JQ1607*
Agrotis segetum NPV	AgseMNPV	P31036
Archips cerasivoranus NPV	ArceNPV	U40834
Attacus ricini NPV	AtriNPV	S68462
Bombyx mori NPV	BmNPV	L33180
Buzura suppressaria SNPV	BusuNPV	X70844
Choristoneura fumiferana NPV	CfMNPV+	
C. fumiferana defective NPV	CFDEF	U78194
Choristoneura rosaceana NPV	ChroNPV	U91940
Ecotropis obliqua SNPV	EcobNPV	U95014
Epiphyas postvittana NPV	EppoNPV	AF061577
Helicoverpa armigera SNPV	HaSNPV	U97657
H. armigera SNPV (South Africa)	HaSNPV-SA	AF157012
H. armigera SNPV (Isareal)	HaSNPV-IS	AJ001917
H. zea SNPV	HzSNPV	Z12117
Hyphantria cunea MNPV	HycuNPC	D14573
Lambdina fiscellaria fisecellaria MNPV	LafiNPV1#	
L. fiscellaria lugubrosa MNPV	LafiNPV2#	
Lymantria dispar MNPV	LdMNPV	M23167
Leucania separata MNPV	LeseNPV	U30302
Lonomia obliqua MNPV	LoobNPV	AF232690
Mamestra brassicae MNPV	MbMNPV	M20927
Malacosoma disstria NPV	MadiNPV	U61732
Mamestra configurata MNPV	MacoNPV	U59461
Malacosoma neustria NPV	ManeNPV	X55658
Orgyia pseudotsugata MNPV	OpMNPV	M14885
Orgyia pseudotsugata SNPV	OpSNPV	M32433
Orgyia anartoides NPV	OranNPV	AF068188
Panolis flammea MNPV	PafiNPV	D00437
Perina nuda MNPV	PenuNPV	U22824
Plusia orichalcea NPV	PlorNPV	AF019882
Rachiplusia ou	RaouNPV	AF068269
Spodoptera exigua MNPV	SeMNPV	X67243
Spodoptera frugiperda MNPV	SpfrNPV	J04333
Spodoptera littoralis MNPV	SpliNPV	D01017.
Spodopetera litura MNPV	SplitNPV1	AF037262
Spodopetera litura MNPV	SplitNPV2	X94437
Trichoplusia ni SNPV	TnSNPV	AF093405
Thysanoplusia orichalcea NPV	ThorNPV	AF169480
Wiseana signata SNPV	WisiNPV	AF016917
Choristoneura fumiferana GV	ChfuGV	U87621
Cydia pomonella GV	CpGV	Y09478
Cryptophlebia leucotreta GV	CrleGV	X79596
Harrisina brillians GV	HabrGV	AF142425
Pieris brassicae GV	PbGV	X02498
Plutella xylostella GV	PxGV	AF270936
Trichoplusia ni GV	TnGV	K02910
Xestia c-nigrum GV	XcGV	U70069

* protein sequence; + from B. M. Arif, personal communication, # from Levin e*t al.*, 1997

It is also interesting to see if any baculovirus gene could be used for quantitative baculovirus taxonomy. So far polyhedrin sequence data of 52 baculovirus species or strains are available (Table 9-4), permitting us to check if they can be used to establish a hierarchical classification of baculoviruses as has been done with potyviruses. For this purpose all polyhedrin amino acid sequences were subjected to pairwise analysis. The frequency distribution of the 1326 pairwise comparisons of baculoviral polyhedrins is showed in Figure 9-4. The lowest level of sequence homology (52-60%) corresponds to genera, indicating the major difference between genus *Nucleopolyhedrovirus* and *Granulovirus*. But we could not find a clear cut off point between species and strain as found for potyviruses (Fig. 9-4). The second level of sequence similarity formed a broad continuum ranging from 78% to 100%. However, one peak located from 78% to 94% might reflect the existence of distinct baculovirus species. The highest level of sequence similarity (98-100%) in polyhedrin corresponds to strains of same baculovirus species, *e.g.* TnGV and CerGV; HaSNPV and HzSNPV; OpMNPV, PenuMNPV and ChroNPV. The other viruses, *e.g.* AcMNPV, AnfaNPV and RaouNPV, which have been classed as different strains of the same virus species, exhibited lower homology than 98% sequence similarity, which might be the case when polyhedrin genes were transferred among baculoviruses later in their evolutionary history. The other genes, such as DNA polymerase and *egt* did not give a clear picture, predominantly due to the lack of sufficient available data later in their evolutionary history. The other genes, such as DNA polymerase and *egt* did not give a clear picture, predominantly due to the lack of sufficient

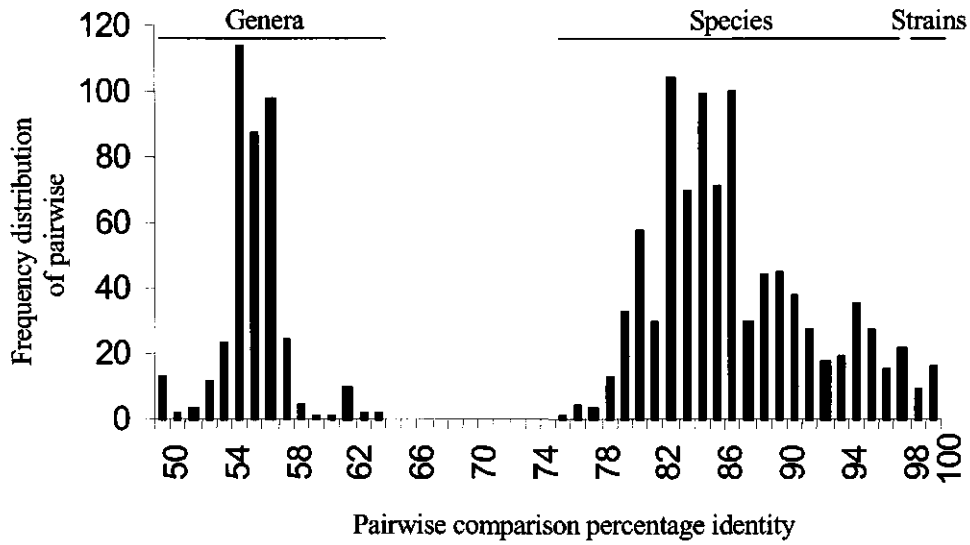


Figure 9-4. Frequency distribution of pairwise amino acid sequence comparisons of 52 polyhedrins/granulins within the family *Baculoviridae*.

available data information (results not shown). With the amount of sequence data rapidly increasing, one gene or part of a gene could be used for baculovirus quantitative taxonomy based on the sequence identity. However, classifying viral gene and genomes should not be confused with classifying viruses. Genome comparisons can not by themselves justify taxonomic placements that would disregard other biotic and phenotypic properties of viruses.

CONSTRUCTION AND EVALUATION OF RECOMBINANT HaSNPVs WITH IMPROVED INSECTICIDAL PROPERTIES

HaSNPV has been developed as a commercial biopesticide to control the cotton bollworm, *H. armigera*, in China. The major limitation to a broader application has been the relatively slow speed of action. To overcome this disadvantage, HaSNPV recombinants have been constructed by deletion of the *egt* gene and/or insertion of an insect-specific scorpion toxin (AaIT) gene (Chapter 8). Such recombinants killed the insect host faster than wild type virus. To test the field efficiency of such recombinants, a plot experiment was conducted with two recombinant viruses, *e.g.* an *egt*$^-$ mutant (HaCXW1) and an *egt*$^-$ and AaIT$^+$ mutant (HaCXW2), on cotton (Sun *et al.*, 2001). Treatments with HaCXW1 and HaCXW2 resulted in 15.3% and 26.3% reduction in ST_{50} values as compared to HaSNPV wt. Feeding reduction by larvae infected with HaCXW1 and HaCXW2 approximated 50% and 63%, respectively, compared to HaSNPV wt. These results provisionally indicate that in a field situation the recombinants are more effective control agents of the cotton bollworm than wild-type HaSNPV.

In most cases baculoviruses have been mass-produced in factories under controlled conditions *in vivo* (Shapiro, 1982, 1986; Shieh, 1989). In China techniques of rearing bollworm on artificial diet and virus production were developed in the late 1970s (Zhang *et al.*, 1983b). The general production protocol for the HaSNPV pesticide consists of insect rearing, viral infection, viral harvest, processing for preparation, qualitative control and package. The critical step is the amplification of polyhedra in the insect.

The development of faster-killing recombinants together with the advances in formulation and application technology has created new market possibilities. However, with the increase of killing speed of the recombinant virus strains, the insect larva infected with those recombinants yielded less polyhedra compared with the larva infected with the wild type (Dupont, 1996; Slavicek *et al.*, 1999). In general, it might be anticipated that the better the insecticidal properties are, provided by a recombinant baculovirus due to faster killing, the less likely it is that the recombinant could be economically produced *in vivo*. In our research we also found that in *H. armigera* larvae *egt*$^-$ and *egt*$^-$AaIT$^+$ HaSNPV replicated to less than 50% of the wild type virus (unpublished data). This makes it difficult to economically produce recombinant virus strains *in vivo*.

One way to solve this problem is to produce such recombinant viruses in insect cell cultures. However, due to the high costs of this method and the lack of stable cell-virus systems available for mass-production, it is difficult at this moment to produce recombinant viral insecticide *in vitro* on an economic basis. An alternative strategy may therefore be the generation of recombinants, in which the insecticidal genes are driven by inducible/suppressible promoters, such as temperature sensitive, antibiotic, metal- or hormone inducible/suppressible promoters. Deletion of *egt* gene may become a common strategy to generate recombinant baculovirus for insect control. To conditionally inactive this gene the *egt* promoter could be replaced by an inducible promoter, *e.g.* the promoter of the *Drosophila melanogaster* metallothionein gene (*Mtn*), which remains inducible in lepidoptertan cells (Lanier *et al.*, 1997; Hegedus *et al.*, 1998). It may be expected that such recombinants could still behave like wild-type viruses when produced in the presence of Cu^{++}, while they would exhibit the *egt*$^-$ phenotype when tested in the field.

A successful approach has been the "Tet" transactivator system (Dupont, patent application.). Here the toxin gene is under control of a promoter that is activated by the tetracycline transactivator protein, which is also expressed by the recombinant baculovirus. During the production, tetracycline is added in the diet and binds to the transactivator protein. As a result the toxin gene transcription and its translation do not occur. The yield of recombinant virus thus will be the same as with wild-type virus. In the field, in the absence of tetracycline, the toxin protein will be produced resulting in faster death of target insects. This system has been successfully used to produce a recombinant HzSNPV with toxin and appeared to be safe and efficacious in the field. It remains to be seen, however, whether such a recombinant will be commercially used and pass the regulatory bodies in view of the mounting public concern on GMOs. However, the application of this new technology could outweigh the environmental consequences of the use of chemical insecticides in cotton growing in China today. *Egt*-deletion mutants (Chapter 8) may be the acceptable choice as such mutants also exist in wild-type baculovirus population (Dai *et al.*, 2000) and there is less trade-off when produced *in vivo* as compared with toxin-producing baculovirus.

REFERENCES

Ahrens, C.H., Rohrmann, F., 1995, Replication of the *Orgyia pseudotsugata* baculovirus DNA: *lef-2* and *ie-1* are essential and *ie-2*, *p34* and *Op-iap* are stimulatory genes. *Virology* 212: 650-662.

Ahrens, C. H., Russell, R., Funk, C. J., Evans, J. T., Harwood, S. H., Rohrmann, G. F. 1997, The sequence of the *Orgyia pseudotsugata* multinucleocapsid polyhedrosis virus genome. *Virology* 229: 381-399.

Albrecht, J. C., Nicholas, J., Biller, D., Cameron, K. R., Biesinger, B., Newman, C., Wittmann, S., Craxton, M. A., Coleman, H., Fleckenstein, B., Honess, R. W. 1992, Primary structure of the herpesvirus saimiri genome. *Journal of Virology* 66: 5047-58.

Altschul, S. F., Gish, W., Miller, W., Meyers, E. W., Lipman, D. J. 1990, Basic local alignment search tool. *Journal of Molecular Biology* 215: 403-410.

Arif, B. M., Doerfler, W. 1984, Identification and location of reiterated sequences in the *Choristoneura fumiferana* NPV genome. *EMBO Journal* 3: 525-529.

Ayres, M. D., Howard, S. C. Kuzio, J., Lopez-Ferber, M., Possee, R. D. 1994, The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. *Virology* 202: 586-605.

Bairoch, A. P., Buchner, P. & Hofmann, K. 1997, The PROSITE database: its status in 1997. *Nucleic Acids Research* 25: 217-221.

Barrett, J. W., Krell, P. J., Arif, B. M., 1995, Characterization, sequencing and phylogeny of the ecdysteroid UDP-glucosyltransferase gene from two distinct nuclear polyhedrosis viruses isolated from *Choristoneura fumiferana*. *Journal of General Virology* 76: 2447-2456.

Barrett, J. W., Brownwright, A. J., Primavera, M. J., Palli, S. R. 1998, Studies of the nucleopolyhedrovirus infection process in insects by using the green fluorescence protein as a reporter. *Journal of Virology* 72: 3377-3382.

Bianchi, F. J. J. A., Snoeijing, I., Van der Werf, W., Mans, R. M. W., Smits, P. H., Vlak, J. M. 2000, Biological activity of SeMNPV, AcMNPV and three AcMNPV deletion mutants against *Spodoptera exigua* larvae (Lepidoptera, Noctuidae). *Journal of Invertebrate Pathology* 75: 28-35.

Bideshi, D. K., Bigot, Y., Federici, B. A. 2000, Molecular characterization and phylogenetic analysis of the *Harrisina brillians* granulovirus granulin gene. *Archives of Virology* 145: 1933-1945.

Birnbaum, M. J., Clem, R. J., Miller, L. K. 1994, An apoptosis-inhibiting gene from a nuclear polyhedrosis virus encoding a polypeptide with Cys/His sequence motifs. *Journal of Virology* 68: 2521-2528.

Bischoff, D. C., Slavicek, J. M. 1999, Impact of deletion of the *Lymantria dispar* nucleopolyhedrovirus PEP gene on viral potency: Expression of the green fluorescent protein prevents larval liquefaction. *Biological Control* 14: 51-59.

Black, B. C., Brennan, L. A., Dierks, P.M., Gard I. E. 1997, Commercialization of baculoviral insecticides. In "*The Baculoviruses*" (Miller L. K., ed). Plenum Press. New York and London. pp 341-387.

Blanchette, M., Kunisawa, T., Sankoff, D. 1999, Gene order breakpoint evidence in animal mitochondrial phylogeny. *Journal of Molecular Evolution* 49: 193-203.

Blissard, G. W. 1996, Baculovirus-insect cell interactions. *Cytotechnology*. 20: 73-93.

Blissard, G. W, Rohrmann G. F. 1990, Baculovirus diversity and molecular biology. *Annual Review of Entomology* 35: 127-155.

Blissard, G., Black, B., Crook, N., Keddie, B. A., Possee, R., Rohrmann, G., Theilmann, D., and Volkman, L. 2000, Baculoviridae. In "*Virus Taxonomy*" (Regenmortel M. H. V., Fauquet, C.M., Bishop, D. H. L., Carstens, E. B., Estes, M. K., Lemon S. M., Maniloff, J., Mayo, M. A., McGeoch, D. J., Pringle, C. R., and R. B. Wickner, eds.). Academic Press, San Diego. pp195-202.

Bonfield, J. K., Smith, K. F., Staden, R. 1995, A new DNA sequence assembly program. *Nucleic Acid Research* 24: 4992-4999.

Bonning, B. C., Hammock, B. D. 1996, Development of recombinant baculoviruses for insect control. *Annual Review of Entomology* 41: 191-210.

Bonning, B. C., Hirst, M., Possee, R. D., Hammock, B. D. 1992, Future development of a recombinant baculovirus insecticide expressing the enzyme juvenile hormone esterase from *Heliothis virescens*. *Insect Biochemistry and Molecular Biology* 22: 453-458.

Bonning, B. C., Hoover, K., Booth, T. F., Duffey, S., Hammock, B. D. 1995, Development of a recombinant baculovirus expressing a modified juvenile hormone esterase with potential for insect control. *Archives of Insect Biochemistry and Physiology* 30: 177-194.

Boyce, F. M., Bucher, N. L. 1995, Baculovirus-mediated gene transfer into mammalian cell. *Proceeding of the National Academy of Sciences of USA* 93: 2348-52.

Brewer, M. J., Rumble, J. T. 1989, Field monitoring for insecticide resistance in beet armyworm (Lepidoptera: Noctuidae). *Journal of Economic. Entomology.* 82: 1520-1526.

Broer, R., Heldens, J. G. M., Strien E. A. ven, Zuidema, D., Vlak, J.M., 1998, Specificity of multiple homologous genomic regions in *Spodoptera exigua* multicapsid nucleopolyhedrovirus DNA replication. *Journal of General Virology* 79: 1563-1572.

Bulach, D. M., Kumar, C. A., Zaia, A., Liang, B., Tribe, D. E. 1999, Group II nucleopolyhedrovirus subgroups revealed by phylogenetic analysis of polyhedrin and DNA polymerase gene sequences. *Journal of Invertebrate Pathology* 73: 59-73.

Carbonell, L. F., Hodge, M. R., Tomaski, M. D., Miller, L. K. 1988, Synthesis of a gene coding for an insect-specific scorpion neurotoxin and attempts to express it using baculovirus vectors. *Gene* 73: 409-418.

CCRI-China Cotton Research Institute. 1983, Cultivation of cotton in China. Shanghai Science and Technology Press, Shanghai. pp 743.

Chao, Y-C., Chen, S.-L., Li, C.-F. 1996, Pest control by fluorescence. *Nature* 380: 386-397.

Chen, C. J.; Thiem, S. M. 19997, Differential infectivity of two *Autographa californica* nucleopolyhedrovirus mutants on three permissive cell lines is the result of *lef*-7 deletion. *Virology* 227:88-95.

Chen, X., Hu, Z, Jehle, J. A., Zhang, Y., Vlak, J. M. 1997a, Analysis of the ecdysteroid UDP-glucosyltransferase gene of *Heliothis armigera* single-nucleocapsid baculovirus. *Virus Genes* 15: 219-225.

Chen, X., Hu, Z. H. Vlak, J. M. 1997b, Nucleotide sequence analysis of the polyhedrin gene of *Heliothis armigera* single nucleocapsid nuclear polyhedrosis virus. *Virologica Sinica* 12: 346-353.

Chen, X., IJkel, W. F. J., Dominy, C., Zanotto, P. de A., Hashimoto, Y., Faktor, O., Hayakawa, T., Wang, C. H., Prekumar A., Mathavan, S., Krell, P. J., Hu, Z., Vlak, J. M. 1999, Identification, sequence analysis and phylogeny of the *lef-2* gene of *Helicoverpa armigera* single-nucleocapsid baculovirus. *Virus Research* 65: 21-32.

Chen, X., Li, M., Sun, X., Arif, B. M., Hu, Z. H., Vlak, J. M. 2000a, Genomic organization of *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus. *Archives of Virology* 145: 2539-2555.

Chen, X., Sun, X., Hu, Z. H., Li, M., O'Reilly, D. R., Zuidema, D., Vlak, J. M. 2000b, Genetic engineering of *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus as an improved bioinsecticide. *Journal of Invertebrate Pathology* 76: 140-146.

Chen, X., IJkel, W. F. J, Tarchini R., Sun, X., Sandbrink H., Wang, H. Peters, S., Zuidema D., Lankhorst, R. K., Vlak, J. M., Hu, Z. 2001, The sequence of the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus genome. *Journal of General Virology* 82: 241-257.

Chisholm, G. E., Henner, D. J. 1988, Multiple early transcripts and splicing of the *Autographa californica* nuclear polyhedrosis virus IE-1 gene. *Journal of Virology* 62: 3193-3200.

Chou, J.M., Huang, C. J., Lo, C. F., Kou, G. H., Wang, C. H. 1996, Characterization of the *Perina nuda* nucleopolyhedrovirus (PenuNPV) polyhedrin gene. *Journal of Invertebrate Pathology* 67: 259-266.

Clarke, E. E., M., Cory, J. S., O'Reilly, D. R. 1996, Characterization of the ecdysteroid UDP-glucosyltransferase gene from the *Mamestra brassicae* nuclear polyhedrosis virus. *Journal of General of Virology* 77: 2865-2871.

Clem, R. J., Miller, L. K. 1994, Control of programmed cell death by the baculovirus genes *p35* and *iap*. *Molecular Cell Biology* 14: 5212-5222.

Cochran, M. A., Faulkner, P. 1983, Location of homologous DNA sequences interspersed at five regions in the baculovirus AcMNPV genome. *Journal of Virology* 45: 961-970.

Condreay, J. P., Witherspoon, S. M., Clay, W. C., Kost, T. A. 1999, Transient and stable gene expression in mammalian cells transduced with a recombinant baculovirus vector. *Proceeding of the National Academy of Sciences of USA* 96: 127-132.

Corsaro, B. B., Fraser, M. J. 1987, Characterization of genotypic and phenotypic variation in plaque-purified strains of HzSNPV Elcar isolate. *Intervirology* 28: 185-198.

Cory, J. S., Hirst, M. L., Williams, T., Hails, R. S. Goulson, D., Green, B. M. 1994, Field trial of a genetically improved baculovirus insecticide. *Nature* 370: 138-140.

Cown, P., Bulach, D., Goode, K., Robertson, A., Tribe, E. 1994, Nucleotide sequence of the polyhedrin gene region of *Helicoverpa zea* single nucleocapsid nuclear polyhedrosis virus: placement of the virus in lepidoptera nuclear polyhedrosis virus group II. *Journal of General Virology* 75: 3211-3218.

Croizier G., Croizier, L., Argaud, O., Poudevigne, D. 1994, Extension of *Autographa californica* nuclear polyhedrosis virus host range by interspecific replacement of a short DNA sequence in the p143 helicase gene. *Proceeding of National Academy Sciences of USA* 91: 48-52.

Crook, N. E. Clem, R. J., Miller, L. K. 1993, An apoptosis-inhibiting baculovirus gene with a zinc finger-like motif. *Journal of Virology* 67: 2168-2174.

Cummings, M. P., Otto S. P., Wakely, J. 1995, Sampling properties of DNA sequence data in phylogenetic analysis. *Molecular Biology and Evolution* 12: 650-656.

Cunningham, J. C. 1995, Baculovirus as microbial insecticides. In "*Novel Approaches to Integrated Pest Control*" (Reuven, R., ed). CRC Press Inc. Boca Raton, Florida, USA, pp261-292.

Cunningham, J. C. 1998. North America. In "*Insect Virus and Pest Management*" (Hunter-Fujita F.R., Entwistle, P. F., Evans H. F., Crook, N. E., ed) Wiley, Chichester, pp 313-331.

Dai, X., Shi, X., Pang, Y., Su, D. 1999. Prevention of baculovirus-induced apoptosis of BTI-Tn-5B1-4 (Hi5) cells by the p35 gene of *Trichoplusia ni* multicapsid nucleopolyhedrovirus. *Journal of General Virology* 80: 1841-1845.

Dai, X., Hajos, J. P., Joosten, N. N., Van Oers, M. M., IJkel, W. F. J., Zuidema, D., Pang, Y., Vlak J. M. 2000, Isolation of a *Spodoptera exigua* baculovirus recombinant with a 10.6 kbp genome deletion that retains biological activity. *Journal of General Virology* 81: 2545-54.

D'Amours, D., Desnoyers, S., d'Silva, I., Poirier, G. G. 1999, Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions. *Biochemical Journal* 342: 249-268.

Dandekar, T., Snel, B., Huynen, M., Bork, P. 1998, Conservation of gene order: a fingerprint

of proteins that physically interact. *Trends Biochemistry Science* 23: 324-328.
Devereux, J., Haeberli, P., Smithies, O. 1984, A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* 12: 387-395.
Domingo, E., Holland, J. J., Biebricher, C., Eigen, M. 1995, Quasispecies: the concept and the world. In "*Molecular Basis of Evolution*". (A. Gibbs, C. Calisher and F. Garcia-Arenal, ed). Cambridge University Press, Cambridge, pp 171-180.
Dreschers, S., Roncarati, R., Knebel-Mörsdorf, D. 2001, Actin rearrangement-inducing factor of baculoviruses is tyrosine phosphorylated and colocalizes to F-actin at the plasma membrane. *Journal of Virology* 75: 3771-3778.
DuPont, 1996, Notification to conduct small-scale field testing of a genetically altered baculovirus, EPA No. 352-NMP-4.
Eason, J.E., Hice, R.H., Johnson, J.J., Federici, B. A. 1998, Effects of substituting granulin or a granulin-polyhedrin chimera for polyhedrin on virion occlusion and polyhedral morphology in *Autographa californica* multinucleocapsid nuclear polyhedrosis virus. *Journal of Virology* 72: 6237-43.
Eldridge, R., O'Reilly D. R., Hammock B. D., Miller, L. K. 1992a, Insecticidal properties of genetically engineered baculoviruses expressing an insect juvenile hormone esterase gene. *Applied and Environmental Microbiology* 58: 1583-1591.
Eldridge, R., O'Reilly D. R., Miller, L. K. 1992b, Efficacy of a baculovirus pesticide expressing an eclosion hormone gene. *Biological Control* 2: 104-110.
Engelhard, E.K., Kam-Morgan, L. N. W., Washburn, J. O., Volkman, L.E. 1994, The insect tracheal system: A conduit for the systemic spread of *Autographa californica* M nuclear polyhedrosis virus. *Proceeding of the National Academy of Sciences of USA* 91: 3224.
Enright, A., Ilipoulos, I., Kyrpides, N., Ouzounis, C. 1999, Protein interaction maps for complete genomes based on gene fusion events. *Nature* 402: 86-90.
Entwistle. P. F. 1998. A world survey of virus control of insect pests. In "*Insect Virus and Pest Management*" (Hunter-Fujita F.R., Entwistle. P. F., Evans H. F., Crook, N. E., eds) Wiley, Chichester, pp 189-200.
Evans, H. F., 1981, Quantitative assessment of the relationship between dosage and response of the nuclear polyhedrosis virus of *Mamestra brassicae*. *Journal of Invertebrate Pathology* 37: 101-109.
Evans, J. T., Leisy, D. J., Rohrmann, G. F. 1997, Characterization of the interaction between the baculovirus replication factors LEF-1 and LEF-2. *Journal of Virology* 71: 3114-3119.
Evans, R. M., Hollenberg, S. M. 1988, Zinic fingers: guilt by association. *Cell* 15:1-3.
Ewing, B., Hillier, L., Wendl, M. C., Green, P. 1998, Basecalling of automated sequencer traces using PHRED. I. Accuracy assessment. *Genome Research* 8: 175-185.
Ewing, B., Green, P. 1998, Basecalling of automated sequencer traces using PHRED. II. Error probabilities. *Genome Research* 8: 186-194.
Fabian, P., Murvai, J., Hatsagi, Z., Vlahovicek, K., Hegyi, H., Pongor, S. 1997, The SBASE protein domain library, release 5.0: a collection of annotated protein sequence segments. *Nucleic Acids Research* 25: 240-243.
Faktor, O., Toister-Achituv, M., Kamensky, B. 1995, Identification and nucleotide sequence of an ecdysteroid UDP-glucoyltransferase gene of *Spodoptera littoralis* multicapsid nuclear polyhedrosis virus. *Virus Genes* 11: 47-52.
FAO/WHO. 1973, The use of the virus for the control of insect pests and disease vectors. *WHO Technical Report No.531*. pp 48.
Federici, B. A. 1986, Ultrastructure of baculovirus. In "*The Biology of Baculoviruses*" (Granados, R.R., and Federici B. A. ed). CRC. pp 61-88.
Federici, B.A. 1999, Naturally occurring baculoviruses for insect pest control. *Methods in Biotechnology* 5: 301-320.

Federici, B. A., Hice, R. H., 1997, Organization and molecular characterization of genes in the polyhedrin region of the *Anagrapha falcifera* multinucleocapsid NPV. *Archives of. Virology* 142: 333-348.

Felsenstein, J. 1978. Case in which parsimony and compatibility methods will be positively misleading. *Systematic Zoology* 27: 401-410.

Figueiredo, E., Muñoz, D., Escribano, A., Mexia, A., Vlak, J. M., Caballero, P. 1999, Biochemical identification and comparative insecticidal activity of nucleopolyhedrovirus isolates pathogenic for *Heliothis armigera* (Lep., Noctuidae) larvae. *Journal of Applied Entomology* 123: 165-169.

Fitt, G. P. 1989, The ecology of *Heliothis* in relation to agroecosystems. *Annual Review of Entomology* 34: 17-52.

Fitz-Gibbon, S. T. and House, C. H. 1999, Whole genome-based phylogenetic analysis of free-living microorganisms. *Nucleic Acids Research* 27: 4218-4222.

Flipsen, J. T. M., van Lent, J.W.M., Goldbach, R. W., Vlak J. M. 1993, Expression of polyhedrin and *p10* in the midgut of AcMNP-infected *Spodoptera exigua* larvae: An immunoelectron microscope investigation. *Journal of Invertebrate Pathology* 56: 17-32.

Flipsen, J. T. M., Mans, R. M. W., Kleefsman, A. W. F., Knebel-Mösrdorf, D., Vlak, J. M. 1995a, Deletion of the baculovirus ecdysteroid UPD-glucosyltransferase gene induces early degeneration of Malpighian tubules in infected insects. *Journal of Virology* 69: 4529-4532.

Flipsen, J. T. M., Martens, J. M. W., Van Oers, M. M., Goldbach, R. W., Vlak J. M., van Lent, J.W.M., 1995b. Passage of *Autographa californica* nuclear polyhedrosis virus through the midgut epithelium of *Spodoptera exigua* larvae. *Virology* 208: 328-335.

Friesen, P. D. 1997, Regulation of Baculovirus early gene expression. In "The *Baculoviruses*" (Miller L. K., ed). Plenum Press, New York/London, pp141-170.

Friesen, P. D., Miller, L. K. 1985. Temporal regulation of baculovirus RNA: overlapping early and late transcripts. *Journal of Virology* 54: 392-400.

Funk, C. J., Braunagel, S. C., Rohrmann, G. F. 1997, Baculovirus structure. In "The *Baculoviruses*" (Miller L. K., ed.). Plenum Press, New York/London, pp7-32.

Garcia–Maruniak, A., Pavan, O. H. O., Maruniak, J. E. 1996, A variable region of *Anticarsia gemmatalis* nuclear polyhedrosis virus contains tandemly repeated DNA sequences. *Virus Research* 41: 123-132.

Gard, I. E. 1997, Field testing a genetically modified virus. In "*Microbial Insecticides: Novelty or Necessity*?", BCPC Symposium No 68, pp 101-114.

Gershburg, E., Stockhom, D., Froy, O., Rashi, S., Gurevitz M., Chejanovsky N. 1998, Baculovirus-mediated expression of a scorpion depressant toxin improves the insecticidal efficacy achieved with excitatory toxin. *FEBS Letters* 422: 132-136.

Gettig, R. R., McCarthy, W. J. 1982, Genotypic variation among wild isolates of *Heliothis* spp nuclear polyhedrosis viruses from different geographical regions. *Virology* 117: 245-252.

Gombart, A. F., Pearson, M. N., Rohrmann G. F., Beaudreau, G. S. 1989, A baculovirus polyhedral envelope-associated protein: genetic location, nucleotide sequence, and immunocytochemical characterization. *Virology* 169: 182-193.

Gomi, S., Majima, K., Maeda, S. 1999, Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus. *Journal of General Virology* 80: 1323-1337.

Gooijer, C. D. de. 1995, Cascades of bioreactors. Ph. D. Thesis, Wageningen Agricultural University, Wageningen, The Netherlands.

Gopalakrishnan, K., Mutukrishnan, S., Kramers, K. J. 1995, Baculovirus-mediated expression of a *Manduca sexta* chitinase gene. *Insect Biochemistry and Molecular Biology* 25: 475-485.

Goto, C., Hayakawa, T., Maeda, S. 1998, Genome organization of the *Xestia c-nigrum*

granulovirus. *Virus Genes* 16: 199-210.

Gould, F., Anderson, A., Jones, A., Sumerford, D., Hechel, D. G., Lopez, J., Micinski, S., Leonard, R., Laster, M. 1997, Initial frequency of alleles for resistance to *Bacillus thuringiensis* toxin in field population of *Heliothis virescens. Proceeding of the National Academy of Sciences of USA* 94: 3519-3523.

Granados, R. R., Lawler, K. A. 1981, In vivo pathway of *Autographa californica* baculovirus invasion and infection. *Virology* 90: 170-174.

Graur, D., Li, W-H. 2000, Genome evolution. In "*Fundamentals of molecular evolution*" (Graur, D. and Li, W-H., eds), Sunderland, Sinauer, pp 367.

Graybeal, A. 1998, It better to add taxa or characters to a difficult phylogenetic problem? *Systematic Biology* 47: 9-17.

Greer, F., Ignoffo, C.M., Anderson, R. F. 1971, The first viral insecticide: a case history. *Chemtechnology* Jun: 324-327.

Guarino, L. A., Summers, M. D. 1986, Interspersed homologous DNA of *Autographa californica* nuclear polyhedrosis virus enhances delayed-early gene expression. *Journal of Virology* 60: 215-223.

Guarino, L. A., Smith, M. 1992, Regulation of delayed-early gene transcription by dual TATA boxes. *Journal of Virology* 66: 3733-3739.

Guarino, L. A., Gonzales, M. A., Summers, M. D. 1986, Complete sequence and enhancer function of the homologous DNA regions of *Autographa californica* nuclear polyhedrosis virus. *Journal of Virology* 60: 224-229.

Guarino, L. A., Xu, B., Dong, W. 1998, A virus-encoded RNA polymerase purified from baculovirus-infected cells. *Journal of Virology* 72: 7985-91.

Hajos, J. P., Vermunt, A. M., Zuidema, D., Kulcsar, P., Varjas, L., de Kort, C. A., Zavodszky, P., Vlak, J. M. 1999, Dissecting insect development: baculovirus-mediated gene silencing in insects. *Insect Molecular Biology* 8: 539-544.

Hamm, J. J. 1982, Relative susceptibility of several noctuid species to a nuclear polyhedrosis virus from *Heliothis armigera. Journal of Invertebrate of Pathology* 39: 255-256.

Hammock, B. D., Bonning, B. C., Possee, R. D., Hanzlik, T. N., Maeda, S. 1990, Expression and effects of juvenile hormone esterase in a baculovirus vector. *Nature* 344: 458-461.

Hashimoto, Y., Hayakawa, T., Ueno, Y., Fujita T., Sano, Y., T. Matsumoto. 2000, Sequence analysis of the *Plutella xylostella* granulovirus genome. *Virology* 275: 358-372.

Hawtin, R. E., Arnold, K. A, Ayres, M. D. Zanotto, P. M., Howard, S. C., Gooday, G. W. Chappll, L. H. Kits, P.A., King, L. A., Possee, R. D. 1995, Identification and preliminary characterization of a chitinase gene in the *Autographa californica* nuclear polyhedrosis virus. *Virology* 212: 673-685.

Hawtin, R. E., Zarkowska, T., Arnold, K. A., Thomas, C. A., Gooday, G. W, King, L. A., Possee, R. D. 1997, Liquefaction of *Autographa californica* nuclear polyhedrovirus-infected insect is dependent of the integrity of virus-encoded chitinase and cathepsin genes *Virology* 238: 243-253.

Hayakawa, T., Ko, R., Okano, K., Seong, S., Goto, C., Maeda, S. 1999, Sequence analysis of the *Xestia c-nigrum* granulovirus genome. *Virology* 262: 277-297.

Hayakawa, T., Rohrmann, G. G., Hashimoto, Y. 2000, Patterns of genome organization and content in Lepidopteran baculovirus. *Virology* 278: 1-12.

Hegedus, D. D., Pfeifer, T.A., Hendry, J., Theilmann, D. A., Thomas A. G. 1998, A series of broad host range shuttle vectors for constitutive and inducible expression of heterologous proteins in insect cell lines. *Gene* 207: 241-249.

Heldens, J. G. M., Yi, L., Zuidema, D., Goldbach, R. W., Vlak, J. M. 1998, A highly conserved genomic region in baculoviruses: Sequence and transcriptional analysis of an 11.3 kbp DNA fragment (mu 46.5-55.1) from the *Spodoptera exigua* multicapsid

nucleopolyhedrovirus. *Virus Research* 55: 187-198.

Hernandez-Crespo, P., Hails, R. S., Sait, S. M., Green, B. M., Carty T. M., Cory J. S. 1999, Response of hosts of varying susceptibility to a recombinant baculovirus insecticide in the field. *Biological Control* 16: 119-127.

Herniou, E. A., Luque, T., Chen, X., Vlak, J. M., Winstanley, D., Cory, J. S, O'Reilly, D. 2001, Baculovirus phylogenies based on gene sequences, gene order and gene content. *Journal of Virology*. In press.

Hill, J. E., Kuzio, J., Faulkner, P. 1995, Identification and characterization of the *v-cath* gene of the baculovirus CfMNPV. *Biochemical and Biophysical Acta*. 1264: 275-278.

Hseih, T. R. 1989, Industrial production of viral pesticides. *Advances in Virus Research* 36: 315-343.

Hofmann, C., Sandig, V., Jennings, G., Rudolph, M., Schlag, P., Strauss, M. 1995, Efficient gene transfer into human hepatocytes by baculovirus vectors. *Proceedings of the National Academy of Sciences of USA* 92: 10099-10103.

Hooft van Iddekinge, B. J. L., Smith, G. E., Summers, M. D. 1983, Nucleotide sequence of the polyhedrin gene of *Autographa californica* nuclear polyhedrosis virus. *Virology* 131:561-565.

Hu, Z. 1998, Characterization of the *Buzura suppressaria* single-nucleocapsid nucleopolyhedrovirus genome: a (phylo)genetic study. PhD thesis, Wageningen University.

Hu Z., Vlak, J. M. 1997. Engineering of biosafe baculoviruses with improved insecticidal properties: Developments and prospects. *Virologica Sinica* 12, 14-25.

Hu, Z., Liu, M.F., Jin, F., Wang, Z. .X, Liu, X. Y., Li, M. J., Liang, B. F., Xie, T. E., 1993, Nucleotide sequence of the *Buzura suppressaria* single nucleocapsid nucleopolyhedrosis virus polyhedrin gene. *Journal of General Virology* 74: 1617-1620.

Hu, Z., Broer, R, Westerlaken, J., Martens, J. W. M., Jin, F., Jehle, J. A., Wang, L. M., Vlak, J. M. 1997, Characterization of the ecdysteroid UDP-glucosyltransferase gene of a single-nucleocapsid nucleopolyhedrovirus of *Buzura suppressaria*. *Virus Research* 47: 91-97.

Hu, Z., Arif, B. M., Jin, F., Martens, J. W. M., Chen, X. W., Sun, J. S., Zuidema, D. Goldbach, R. W., Vlak, J. M. 1998, Distinct gene arrangement in the *Buzura suppressaria* single-nucleocapsid nucleopolyhedrovirus genome. *Journal of General Virology* 79: 2841-2851.

Hu, Z., Luijckx, T., van Dinten, L. C., van Oers, M. M., Hajos, J. P., Bianchi, F. J., van Lent, J. W., Zuidema, D., Vlak, J. M. 1999, Specificity of polyhedrin in the generation of baculovirus occlusion bodies. *Journal of General Virology*. 80: 1045-1053.

Huber, J. 1986. Use of baculoviruses in pest management programs. In "*Biology of Baculoviruses*" (R. R. Granados and B. A. Federici. eds), Vol. 2, CRC Press, Boca Raton, FL. pp. 181-202.

Hughes, P. R., Gettig, R. R., McCarthy, W. J. 1983, Comparison of the time-mortality response of *Heliothis zea* to 14 isolates of *Heliothis* nuclear polyhedrosis virus. *Journal of Invertebrate Pathology* 41: 256-261.

Hughes, P. R., van Beek, N. A. M., Wood, H. A. 1986, A modified droplet feeding method for rapid assay of *Bacillus thuringiensis* and baculovirus in noctuid larvae. *Journal of Invertebrate Pathology* 58: 187-192.

Hughes, P. R., Wood, H. A., Breen, J. P., Simpson, S. F., Duggan, A. G., Dybas, J. A. 1997, Enhanced bioactivity of recombinant baculovirus expressing insect-specific spider toxins in lepidopteran crop insects. *Journal of Invertebrate Pathology* 69: 112-118.

Huynen, M., Snel, B., Lathe III, W., Bork, P. 2000, Predicting protein function by genomic content: quantitative evaluation and qualitative inferences. *Genome Research* 10: 1204-1210.

Huynen, M. A., Bork, P. 1998, Measuring genome evolution. *Proceedings of the National*

Academy of Sciences of USA 95: 5849-5856.

HZNU- Lab of Insect Virus of Huazhong Normal University, 1978, The histopathology of *Heliothis armigera* caused by a nuclear polyhedrosis virus VHA-273. *Acta Entomologica Sinica* 21:133-136.

Ignoffo, C. M. 1965a, The nuclear-polyhedrosis virus of *Heliothis zea* (Boddie) and *Heliothis virescens* (Fabricius), I. Virus propagation and its virulence. *Journal of Invertebrate Pathology* 7: 217-226.

Ignoffo, C. M. 1965b, The nuclear-polyhedrosis virus of *Heliothis zea* (Boddie) and *Heliothis virescens* (Fabricius), II. Biology and propagation of diet-reared *Heliothis*. *Journal of Invertebrate Pathology* 7: 209-216.

Ignoffo, C. M. 1965c, The nuclear-polyhedrosis virus of *Heliothis zea* (Boddie) and *Heliothis virescens* (Fabricius), III. Effectiveness of the virus against field populations of *Heliothis* on cotton, corn, grain and sorghum. *Journal of Invertebrate Pathology* 7: 209-216.

Ignoffo, C. M. 1973, Development of a viral insecticide: concept to commercialization. *Experimental Parasitology* 33: 380-406.

Ignoffo, C. M. 1981, Living microbial insecticides. In: "*Essays in Applied Microbiology* " (Norris, J. R. and Richmond, M. H., eds). Wiley, New York, pp 2-31.

Ignoffo, C. M., Couch, T. L. 1981, The nucleopolyhedrosis virus of *Heliothis* spp. as a microbial insecticide. In "M*icrobial control of Pests and Plants diseases 1970-1980*" (H. D. Burges, ed.). London: Academic Press, pp 330-363.

IJkel, W. F. J. 2001, The genome of *Spodoptera exigua* multicapsid nucleopolyhedrovirus: a study on unique features. PhD thesis. Wageningen University. pp 59-77

IJkel, W. F. J., Van Strien, E. A., Heldens, J. G. M., Broer, R., Zuidema, D., Goldbach, R. W., Vlak, J. M. 1999, Sequence and organization of the *Spodoptera exigua* multicapsid nucleopolyhedrovirus genome. *Journal of General Virology* 80: 3289-3304.

IJkel, W. F. J., Westenberg, M., Goldbach, R. W., Blissard, G. W., Vlak, J. M., Zuidema, D. 2000, A novel baculovirus envelope fusion protein with a proprotein convertase cleavage site. *Virology*, 275: 30-41.

IJkel, W. F. J., Lebbink, R. J., Op den Brouw M. L., Goldbach, R. W., Vlak, J. M., Zuidema, D. 2001, Identification of a novel occlusion derived virus-specific protein in *Spodoptera exigua* multicapsid nucleopolyhedrovirus. *Virology* 284: 170-181.

James, C. 2000, Global review of commercialized transgenic crops: 2000. *ISAAA Briefs* No: 21: Preview. ISAAA, Ithaca, NY.

Jarvis, D. L., Bohlmeyer, D. A., Garcia, A., 1992. Enhancement of polyhedrin nuclear localization during baculovirus infection. *Journal of Virology* 66: 6903-6911.

Jarvis, D. L., Reilly, L. M., Hoover, K., Schultz, C., Hammock, B. D. 1996, Construction and characterization of immediate early baculovirus pesticides. *Biological Control* 7: 228-235.

Jehle, J. A., Linden, I. F. A. van der, Backhaus, H., Vlak, J. M. 1997, Identification and sequence analysis of the integration site of transposon TCp3.2 in the genome of *Cydia pomonella* granulovirus. *Virus Research* 50: 151-157.

Jenner, R. G., Alba, M.M., Boshoff, C., Kellam. P. 2001, Kaposi's sarcoma-associated herpsvirus latent and lytic gene expression as revealed by DNA array. *Journal of Virology* 75: 891-902.

Jiang, Z., Shan, H., Xu, Y., He, J., Luo, Y. 1978, Study on *Heliothis armigera* nuclear polyhedrosis virus (I): Bioassay of HaSNPV. *Journal of Fudan University* 4: 85-92.

Jones, K. A. 1994, Use of baculovirus for cotton pest control. In: "*Insect Pest of Cotton*" (Matthews G. A. & Tunstall, J. P), Wallingford, UK: CAB International, pp477-504.

JSM & CCNU — Jingzhou Station of Microbiology, and Bio-control Research Group of Central China Normal University, 1975, Primary test of *Heliothis armigera* nuclear polyhedrosis virus to control cotton bollworm. *Hubei Agricultural Science and Technology*

12: 26-29.
Kang, W., Suzuki, M., Zemskov, E., Okano, K., Maeda, S. 1999, Characterization of baculovirus repeated open reading frames (*bro*) in *Bombyx mori* nucleopolyhedrovirus. *Journal of Virology* 73: 10339-10345.
Kelly, D. C., Brown, D. A., Robertson, J. S., Harrap, K. A. 1980, Biochemical, biophysical and serological properties of two singly enveloped nuclear polyhedrosis viruses from *Heliothis armigera* and *Heliothis zea*. *Microbiologica (Bologna)* 3: 319-331.
Ketunuti, U., Prathomrut, S. 1989, Cotton bollworm larvae control by *Heliothis armigera* nuclear polyhedrosis virus. In: "*Abstracts of the First Asia-Pacific Conference of Entomology*", 8-13 November 1898, Chang Mai, Thailand. The secretariat APEC. Bangkok, pp 16.
Kim, D., Weaver, R. F. 1993, Transcription mapping and functional analysis of the protein tyrosine/serine phosphatase (PTPase) gene of the *Autographa californica* nuclear polyhedrosis virus. *Virology* 195: 587-595.
King, A. B. S. 1994, *Heliothis/Helicoverpa armigera* (Lepidoptera: Noctuidae). In: "*Insect Pests of Cotton*" (Matthews G. A. and Tunstall, J. P, eds), Wallingford, UK: CAB International, pp 445-467.
King, E. G., Coleman, R. J. 1989, Potential for biological control of *Heliothis* species. *Annual Review of Entomology* 34: 53-75.
Kitts P. A., Ayres, M. D., Possee, R. D. 1990, Linearization of baculovirus DNA enhances the recovery of recombinant virus expression vectors. *Nucleic Acids Research.* 18: 5667-5672.
King L. A., Possee, R. D. 1992, The baculovirus expression system: a laboratory guide. Chapman and Hall, London, pp 180.
Knell, J. D., Summers, M. D. 1984, A physical map of the *Heliothis zea* SNPV genome. *Journal of General Virology* 65: 445-450.
Kolodyn-Hirsh, D., Dimmock, M. B. 1996, Commercial development and use of Spod-X^R: a wild type baculovirus insecticide for beet armyworm. Abstract Book of the 29th Annual Meeting of the Society for Invertebrate Pathology, Cordoba, Spain, p43.
Kondo, A., Yamamoto, M., 1998, Some properties of a nucleopolyhedrovirus isolated from diamondback moth larvae in the Shiga prefecture. *Bulletin of the Shiga Agricultural Experiment Station.* 39: 29-34.
Kondrashov, A. S. 1999, Comparative genomics and evolutionary biology. *Current Opinion in Genetics and Development* 9: 624-629.
Kool, M., Ahrens, C. H., Goldbach, R. W., Rohrmann, G. F., Vlak, J. M. 1994, Identification of genes involved in DNA replication of the *Autographa californica* baculovirus. *Proceedings of the National Academy of Sciences of USA* 91: 11212-11216.
Kool, M., Ahrens, C. H., Vlak, J. M., Rohrmann, G. F. 1995, Replication of baculovirus DNA. *Journal of General Virology* 76: 2103-2118.
Kornberg, A., Baker, T. A. 1992, DNA replication. W. H. Freeman & Co, New York.
Korst, T. A., Condreay, J. P. 1999, Recombinant baculoviruses as expression vectors for insect and mammalian cells. *Current Opinion in Biotechnology* 10: 428-33.
Korth, K. L., Levings, C. S. III. 1993, Baculovirus expression of the maize mitochondrial protein URF 13 confers insecticidal activity in cell cultures and larvae. *Proceedings of the National Academy of Sciences of USA* 90: 3388-3392.
Kunimi, Y., Fuxa, J. R., Hammock, B D. 1996, Comparison of wild type and genetically engineered nuclear polyhedrosis viruses of *Autographa californica* for mortality, virus replication and polyhedra production in *Trichoplusia ni* in larvae. *Entomologia Experimentaliset Applicata* 81: 251-257.
Kunimi, Y., Fuxa, J. R., Hammock, B D. 1997, Survival times and lethal doses of wild and recombinant *Autographa californica* nuclear polyhedrosis viruses in different instars of.

Pseudoplusia includens. *Biological Control* 9: 129-135.

Kuzio, J., Pearson, M. N., Harwood, S. H., Funk, C. J., Evans, J. T., Slavicek, J. M., Rohrmann, G. F. 1999, Sequence and analysis of the genome of a baculovirus pathogenic for *Lymantria dispar*. *Virology* 253: 17-34.

Lanier, L. M., Storm, K., Shafaie, A., Volkman L. E. 1997, Copper treatment increases recombinant baculovirus production and polyhedrin and p10 expression. *BioTechniques* 23: 728-735.

Le, T. H., Wu, T, Robertson, A., Bulach, D., Cowan, P., Goodge, K., Tribe, D. 1997, Genetically variable triplet repeats in a RING-finger ORF of *Helicoverpa* species baculoviruses. *Virus Research* 49: 67-77.

Lee, Y., Qu, X., Chen, W., Poloumienko A., MacAfee, N., Morin, B., Lucarotti, C., Krause, M. 1997, Insecticidal activity of a recombinant baculovirus containing an anti-sense *c-myc* fragment. *Journal of General Virology* 78: 273-381.

Leisy, D., Nesson, M., Pearson, M., Rohrmann, G. F., Beaudreau G. S. 1986 Location and nucleotide sequence of the *Orgyia pseudotsugata* single nucleocapsid nuclear polyhedrosis virus polyhedrin gene. *Journal of General Virology* 67:1073-1079.

Levin, D. B., Laitinen, A. M., Clarke, T., Lucarotti, C, J., Morin, B., Otvos, I. S.1997, Characterization of nuclear polyhedrosis virus from three subspecies of *Lambdina fiscellaria*. *Journal of Invertebrate Pathology* 69: 125-134.

Liu, Y. B., Tabashnik, B. E., Dennehy, T. J., Patin, A. L., Bartlett, A. C. 1999, Development time and resistance to Bt crops. *Nature* 400: 519.

Li, X.; Pang, A.; Lauzon, H. A. M.; Sohi, S. S.; Arif, B. M. 1997, The gene encoding the capsid protein P82 of the *Choristoneura fumiferana* multicapsid nucleopolyhedrovirus: sequencing, transcription and characterization by immunoblot analysis. *Journal of General Virology* 78: 2665-2673.

Lockhart, D. J., Winzeler, E. A. 2000, Genomics, gene expression, and DNA arrays. *Nature* 405: 827-836.

Lu, A., Carstens, E. B. 1992, Nucleotide sequence and transcriptional analysis of the p80 gene of *Autographa californica* nuclear polyhedrosis virus: a homologue of the *Orgyia pseudotsugata* nuclear polyhedrosis virus capsid-associated gene. *Virology* 190: 201-209.

Lu, A., Miller, L. K. 1995, The roles of eighteen baculovirus late expression factor genes in transcription and DNA replication. *Journal of Virology* 69: 975-982.

Lu, A., Miller, L. K. 1996, Species-specific effects of the *hcf-1* gene on baculovirus virulence. *Journal of Virology* 70: 5123-5130.

Lu, A., Krell, P. J., Vlak, J. M., Rohrmann, G. F. 1997, Baculovirus DNA replication. in "*The Baculoviruses*". (Miller, L. K., ed.), Plenum Press, New York, pp171-176.

Lu, A., Miller, L. K. 1997, Regulation of baculovirus late and very late gene expression. in "*The Baculoviruses*". (Miller, L. K., ed), Plenum Press, New York, pp177-192.

Lu, A., Seshagiri, S., Miller, L. K. 1996. Signal sequence and promoter effects on the efficacy of toxin-expressing baculoviruses as biopesticides. *Biological. Control* 7: 320-332.

Lua, L. H., Reid, S. 2000, Virus morphogenesis of *Helicoverpa armigera* nucleopolyhedrovirus in *Helicoverpa zea* serum-free suspension culture. *Journal of General Virology* 81: 2531-2543.

Lübbert, H., Doerfler, W. 1984, Transcription of overlapping sets of RNA from the genome of *Autographa californica* nuclear polyhedrosis virus: A novel method for mapping RNAs. *Journal of Virology* 52: 255-265.

Ma, S. -W., Corsaro, B. G., Klebba, P. E., Fraser, M. J. 1993, Cloning and sequence analysis of a p40 structural protein of *Helicoverpa zea* nuclear polyhedrosis virus. *Virology* 192: 224-233.

Ma, P. K., Davis, T, R., Wood, H. A., Knipple, D. C., Roelofs, W. L. 1998, Baculovirus

expression of an insect gene that encodes multiple neuropeptides. *Insect Biochemistry and Molecular Biology*. 28: 239-249.

Maeda, S. 1989, Increased insecticidal effect by a recombinant baculovirus carrying a synthetic diuretic hormone gene. *Biochemical and Biophysical Research Communication* 165: 1177-1183.

Maeda, S., Kamita, S. G., Kondo, A. 1993, Host range expansion of *Autographa californica* nuclear polyhedrosis virus (NPV) following recombination of a 0.6-kilobase-pair DNA fragment originating from *Bombyx mori* NPV. *Journal of Virology* 67: 6234-6238.

Maeda, S., Volrath, S. L., Hanzlik, T. N., Harper, S. A., Majima, K., Maddox, D. W., Hammock, B. D., and Fowler, E. 1991, Insecticidal effects of an insect-specific neurotoxin expressed by a recombinant baculovirus. *Virology* 184: 777-780.

Marcotte, E. M., Pellegrni, M., Ng, H., Rice, W. D., Yeates, T. O., Eisenberg, D., 1999, Detecting protein function and protein-protein interactions from genome sequences. *Science* 285: 751-753.

Martens, J. W. M., Honee, G, Zuidema, D., van Lent J. W. M., Visser, B., Vlak, J. M. 1990, Insecticidal activity of a bacterial crystal protein expressed by a recombinant baculovirus in insect cells. *Applied and Environmental Microbiology* 56: 2764-2770.

Martens, J. W. M., Knoester, M., Weijts, F., Groffen, S. J. A., Hu, Z. H., Bosch, D., Vlak, J. M. 1995. Characterization of baculovirus insecticides expressing tailored *Bacillus thuringiensis* Cry1A(b) crystal proteins. *Journal of Invertebrate Pathology* 66: 249-257.

Martens, J. W. M., Van Oers, M. M., Van de Bilt, B. D., Oudshoorn, P., Vlak, J. M. 1995, Development of a baculovirus vector that facilitates the generation of p10-based recombinants. *Journal of Virological Methods* 52: 15-19.

McCaffery, A. R., Walker, A. W. 1991, Insecticide resistance in the bollworm, *Helicoverpa armigera* from Indonesia. *Pesticide Science* 32: 85-90.

McCarthy, W. J., Mercer, W. E., Murphy, T. F. 1978, Characterization of the DNA from four *Heliothis* nuclear polyhedrosis virus isolates. *Virology* 90: 374-378.

McCutchen, B. F., Choudary, P. V., Crenshaw, R., Maddox, D. W., Kamita, S. G., Palekar, N., Volrath, S., Fowler, E., Hammock, B. D., Maeda, S. 1991, Development of a recombinant baculovirus expressing an insect selective neurotoxin, potential for pest control. *Biotechnology* 9: 848-852.

McIntosh, A. H., Ignoffo, C. M. 1983, Characterization of 5 cell lines established from species of *Heliothis*. *Applied Entomology and Zoology* 18: 262-269.

McKinley, D. J. 1971, Nuclear polyhedrosis virus of the cotton bollworm in Central Africa. *Cotton Growing Review* 48: 297-303.

Merryweather, A. T., Weyer, U., Harris, M, P.G., Hirst, M., Booth, T. F., Possee, R. D. 1990, Construction of genetically engineered baculovirus insecticides containing the *Bacillus thuringiensis* subsp. Kurstaki HD37 delta endotoxin. *Journal of General Virology* 71: 1535-1544.

Merrington, C. L., Kitts, P. A., King, L. A., Possee, R. D. 1996, An *Autographa californica* nucleopolyhedrovirus *lef-2* mutant: Consequences for DNA replication and very late gene expression. *Virology* 217: 338-348.

Miller, L. K. 1995, Genetically engineered insect virus pesticides: present and future. *Journal of Invertebrate Pathology* 65: 211-216.

Miller, L.K. (ed.), 1997. The baculoviruses. Plenum Press, New York & London, pp.1-32.

Mikhailov, V. S., Mikhailova, A. L., Iwanaga, M., Gomi, S., Maeda, S. 1998, *Bombyx mori* nucleopolyhedrovirus encodes a DNA-binding protein capable of destabilizing duplex DNA. *Journal of Virology* 72: 3107-3116.

Mitter, A. Mitter, C. Cegier, J. 2000, More taxa or more characters revisited: combining data from nuclear protein encoding genes for phylogenetic analysis of Noctuidae. *Systematic*

Biology 49: 202-224.
Mohr, S., Rihn, B. 2001, Profiling gene expression in human mesothelioma cells using DNA microarray and high-density filter array technologies. *Bulletin du Cancer* 88: 305-313.
Montague, M. G., Hutchison III, C. A. 2000, Gene content phylogeny of herpesviruses. *Proceedings of National Academy of Sciences of USA* 97: 5334-5339.
Moscardi, F. 1999, Assessment of the application of baculoviruses for control of lepidoptera. *Annual Review of Entomology* 44: 257-289.
Moser, B. A., Becnel, J. J., White, S. E., Afonso, C., Kutish, G., Shanker, S., Almira, E. 2001, Morphological and molecular evidence that *Culex nigripalpus* baculovirus is an unusual member of the family Baculoviridae. *Journal of General Virology* 82: 283-297.
Müller, R., Pearson, M. N. Russell, R. L. Q., Rohrmann, G. F. 1990, A capsid-associated protein of the multicapsid nuclear polyhedrosis virus of *Orgyia pseudostugata*: genetic location, sequence, transcription mapping, and immunocytochemical characterization. *Virology* 176: 133-144.
Muñoz, D., Castillejo, Caballero, P. 1998, Naturally occurring deletion mutants are parasitic genotypes in a wild-type nucleopolyhedrovirus population of *Spodoptera exigua*. *Applied and Environmental Microbiology* 64: 4372-4377.
Murphy, F. A., Fauquet, C. M., Bishop, D. H. L., Ghabrial, S. A., Jarvis, A. W., Martelli, G. P., Mayo, M. A., Summers, M. D. 1995, Virus Taxonomy. In *Sixth Report of the International Committee on Taxonomy of Viruses,* Springer-Verlag, Wien/New York, pp104-113.
Nielsen, H., Engelbrecht, J., Brunak, S., Von Heijne, G. 1997, Identification of prokaryotic and eukaryotic signal peptides and their predicted cleavage sites. *Protein Engineering* 10: 1-6.
Okano, K., Mikhailov, V. S., Maeda, S. 1999, Colocalization of baculovirus IE-1 and two DNA-binding proteins, DBP and LEF-3, to viral replication factories. *Journal of Virology* 73: 110-119.
Oomens, A. G. P., Blissard, G. W. 1999, Requirement for Gp64 to drive efficient budding of *Autographa californica* multicapsid nucleopolyhedrovirus. *Virology* 254: 297-314.
O'Reilly, D. R. 1995, Baculovirus-encoded ecdysteroid UDP-glucosyltransferases. *Insect Biochemistry and Molecular Biology* 25, 541-550.
O'Reilly, D. R. 1997, Auxiliary genes of baculoviruses. In "*The Baculoviruses*" (Miller, L. K.ed), Plenum Press, New York. pp 267-300.
O'Reilly, D. R., Miller L. K. 1989, A baculovirus blocks molting by producing ecdysteroid UDP-glucosyltransferase. *Science* 245: 1110-1112.
O'Reilly, D. R., Miller, L. K. 1990. Regulation of expression of a baculovirus ecdysteroid UDP-glucosyltransferase gene. *Journal of Virology* 64: 1321-1328.
O'Reilly, D. R., Miller, L. K. 1991, Improvement of a baculovirus pesticide by deletion of the *egt* gene. *Bio/Technol.* 9: 1086-1089.
O'Reilly, D. R., Brown, M. R., Miller, L. K. 1992, Alteration of ecdysteroid metabolism due to baculovirus infection of the fall armyworm *Spodoptera frugiperda*, Host ecdysteroids are conjugated with galactose. *Insect Biochemistry and Molecular Biology* 22: 313-320.
O'Reilly, D. R., Kelly, T. K., Masler, E. P., Thyafaraja, B. S. Robson, R. M., Shaw, R. C., Miller, L. K. 1995, Overexpression of *Bombyx mori* prothoracicotropic hormone using baculovirus vectors, *Insect Biochemistry and Molecular Biology* 22: 313-320.
Passarelli, L., Miller, L. K. 1993, Three baculovirus genes involved in late and very late gene expression: *ie-1, ie-n* and *lef-2. Journal of Virology* 67: 2149-2158.
Payne, R. W., Lane, P. W., Digby, P. G. N., Harding, S. A., Leech, P. K., Morgan, G. W., Todd, A. D., Thompson, R., Tunnicliffe Wilson, G., Welham, S. J., and White, R. P. 1993. Genstat 5 release 3 reference manual. Clarendon, Oxford.
Pearson, W. R. 1990, Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods in Enzymology* 183: 63-98.

Pearson, M. N., Bjornson, R. M., Pearson, G. D., Rohrmann, G. F. 1992, The *Autographa californica* baculovirus genome: evidence of multiple replication origins. *Science* 257: 1382-1384.

Pearson, M. N., Groten, C., Rohrmann, G. F., 2000, Identification of the *Lymantria dispar* nucleopolyhedrovirus envelope fusion protein provides evidence for a phylogenetic division of the Baculoviridae. *Journal of Virology*. 74: 6126-6131.

Pellegrini, M., Marcotte E. M., Thompson, J. M., Eisenberg, D., Yeats T. O. 1999, Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proceedings of the National Academy of Sciences of the United States of America* 96: 4285-4288.

Popham, H. J. R., Li, Y., Miller, L. K. 1997, Genetic improvement of *Helicoverpa zea* nuclear polyhedrosis virus as a biopesticide. *Biological Control* 10: 83-91.

Popham, H. J. R., Pellock, B. J., Robson, M., Dierks, P. M., Miller, L. K. 1998, Characterization of a variant of *Autographa californica* nuclear polyhedrosis virus with a nonfunctional ORF603. *Biological Control* 12: 223-230.

Possee, R. D. 1997, Baculoviruses as expression vectors. *Current Opinions in Biotechnology*. 8: 569-572.

Possee, R. D., Cayley, P. J., Cory, J. S., Bishop, D. H. L. 1993, Genetically engineering viral insecticides, New insecticides with improved phenotype. *Pesticide Science* 39: 109-115.

Possee, R. D., Rohrmann, G. F. 1997, Homologous regions. In "*The Baculoviruses*" (L. K. Miller, ed). Plenum Press, New York, pp 129-133.

Prikhod'ko, G. G., Robson, M., Warmke, J., Cohen, C. J., Smith, M. M., Wang, P., Warren, V., Kaczorowski, G., Van der Ploeg, L. H. T., Miller, L. K. 1996, Properties of three baculoviruses - expressing genes that encode insect-selective toxins: u-Aga-IV, As II, and Sh I. *Biological Control* 7, 236-244.

Reiländer, H., Haase, W., Maul, G. 1996, Functional expression of the *Aequorea victoria* green fluorescent protein in insect cells using baculovirus expression system. *Biochemical and Biophysical Research Communications* 29: 14-20.

Ribeiro, B. M., Crook, N. E. 1993, Expression f full-length and truncated forms of crystal protein genes from *Bacillus thuringiensis* subsp. Kurstaki in a baculovirus and pathogenicity of the recombinant viruses. *Journal of Invertebrate of Pathology* 62: 121-130.

Riegel, C. I., Lanner-Herrera, C., Slavicek, J. M. 1994, Identification and characterization of the ecdysteroid UDP-glucosyltransferase gene of the *Lymantria dispar* multinucleocapsid nuclear polyhedrosis virus. *Journal of General Virology 75*: 829-838.

Rohrmann, G. F. 1986, Polyhedrin structure. *Journal of General Virology* 67: 1499-1513.

Rohrmann, G. F., Pearson, M. N., Bailey, T. J., Becker, R. R., Beaudreau, G. S. 1981, N-terminal polyhedrin sequence and occluded baculovirus evolution. *Journal of Biology and Evolution* 17: 329-333.

Roncarati, R., Knebel-Mörsdorf, D. 1997, Identification of the early actin-rearrangement-inducing factor gene, arif-1, from *Autographa californica* multicapsid nuclear polyhedrosis virus. *Journal of Virology* 71: 7933-7941.

Rost, B., Sander, C. 1993, Prediction of protein secondary structure at better than 70% accuracy. *Journal of Molecular Biology* 232: 584-599.

Russell, R. M., Robertson, J. L., Savin, N. E. 1977. POLO: A new computer program for probit analysis. *Bulletin of the Entomological Society of America* 23: 209-213.

Sambrook, J., Fritsch, E. F., Maniatis, T. 1989, Molecular Cloning: A laboratory Manual, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sanger, F., Nickien, S., Coulson A. R. 1977, DNA sequencing with chain-terminating inhibityors. *Proceedings of the National. Academy of Sciences of USA* 74: 5463-5467.

Sarkis, C., Serguera, C. Peters, S., Buchet, D., Ridet, J.-L., Edelman, L., Mallet, J. 2000, Efficient transduction of neural cells in vitro and in vivo by a baculovirus-derived vector. *Proceedings of the National. Academy of Sciences of USA* 97: 14638-14643.

Shapiro, M. 1982, *In vivo* mass production of insect viruses for use as pesticides. In "*Microbial and Viral Pesticides*" (Kurstak, E., ed). CRC Press, Boca Raton.

Shapiro, M. 1986, *In vivo* production of baculoviruses. In "*Biology of Baculoviruses, vol II Practical Application for Insect Control*" (Granados, R. R. and Federici, B. A. eds.). Marcel Dekker, New York, pp31-61.

Sharp, P. M. 1991, Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. *Journal of molecular Evolution* 33: 23-33.

Shelton, A. M., Tang, J. D., Roush, R. T., Metz, T. D., Earle, E. D. 2000, Field tests on managing resistance to Bt-engineered plants. *Nature Biotechnology* 18: 339-342.

Shieh, T. R. 1989, Industrial production of viral pesticides. *Advances in Virus Research* 36: 315-343.

Shippam, C., Wu, X., Stewart, S., Theilmann, D. A. 1997, Characterization of a unique OpMNPV-specific early gene not required for viral infection in tissue culture. *Virology* 227: 447-459.

Shukla, D. D., Ward, C. W., Brunt, A. A. 1994, The Potyvirus. CAB International Wallingford. pp 178-208.

Slack, J. M., Kuzio, J., Faulkner, P. 1995, Characterization of *v-cath*, a cathepsin L-like proteinase expressed by baculovirus *Autographa californica* multiple nuclear polyhedrosis virus. *Journal of General Virology* 76: 1091-1098.

Slavicek, J. M., Popham, H. J. R., Riegel, C. I. 1999, Deletion of the *Lymantria dispar* multicapsid nucleopolyhedrovirus ecdysteroid UDP-glucosyl transferase gene enhances viral killing speed in the last instar of the gypsy moth. *Biological Control* 16: 91-103.

Smith, G. E., Summers, M. D. Fraser M. J. 1983. Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Molecular and Cellular Biology* 3: 2156-2165.

Smith, I. R., Crook, N. E. 1988, *In vivo* isolation of baculovirus genotypes. *Virology* 166: 240-244.

Smith, I.R.L., Goodale, C. 1998, Sequence and *in vivo* transcription of the *Lacanobia oleracea* granulovirus *egt* gene. *Journal of General Virology* 79: 405-413.

Smits, P. H., Vlak, J. M. 1994, Registration of the first viral insecticide in the Netherlands: the development of Spod-X, based on *Spodoptera exigua* nuclear polyhedrosis virus. *Mededelingen van de Faculteit Landbouwwetenschappen van de Rijksuniversiteit Gent, Belgium* 59: 385-395.

Smits, P. H. van de Vrie, M., Vlak, J. M. 1988, Nuclear polyhedrosis virus control of *Spodoptera exigua* on glasshouse crops. *Entomologia Experimentaliset Applicata.* 43: 73-80.

Snel, B., Bork, P., Huynen, M. A. 1999, Genome phylogeny based on gene context. *Nature Genetics* 21: 108-110.

Southern, E. 1975, Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* 98: 503-517.

Sriram, S., Gopinathan, K. P. 1998, The potential role of a late gene expression factor, *lef2*, from *Bombyx mori* nuclear polyhedrosis virus in very late gene transcription and DNA replication. *Virology* 251, 108-122.

Stewart, L. M., Hirst, M., Ferber, M. L., Merryweather, A. T., Cayley, P. J., Possee, R. D. 1991, Construction of an improved baculovirus insecticide containing an insect-specific toxin gene. *Nature* 352: 85-88.

Sun, S., Chen, X., Zhang, Z., Wang, H., Bianchi, F. J. J. A., Peng, H., Vlak, J. M., Hu, Z.H. 2001. Field evaluation of genetically modified *Helicoverpa armigera* nucleopolyhedrovirus on cotton. *Journal of Invertebrate Pathology*, Accepted.

Sun, X. L., Zhang, G. Y., Zhang, Z. X., Hu, Z. H., Vlak, J. M., Arif, B. M. 1998, *In vivo* cloning of *Helicoverpa armigera* single nucleocapsid nuclear polyhedrosis virus genotypes. *Virologica Sinica* 13: 83-88.

Sun X., Zhang, G., 1994, A comparison of four wild isolates of *Heliothis* nuclear polyhedrosis virus. *Virologica Sinica* 9:309-318.

Swofford, D.L., 1993. PAUP: Phylogenetic analysis using parsimony. Version 3.1. Computer program distributed by the Illinois Natural History Survey, Champaign, IL.

Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperir, M. Y. Fedorova, N. D., Koonin, E. V. 2001, The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Research* 29: 22-28.

Teakle, R. E. 1973, Nuclear-polyhedrosis virus from *Heliothis punctigera* Wallengren (Lepidoptera: Nodtuidae). *Queensland Journal Agricultural and Animal Sciences* 30: 161-177.

Tekaia, F., Lazcano, A., Dujon, B. 1999, The genomic tree as revealed from whole proteome comparisons. *Genome Research* 9: 550-557.

Thiem, S. M. 1997, Prospects for altering host range for baculovirus bioinsecticides. *Current Opinion of Biotechnology* 8:317-322.

Thiem, S. M., Du, X., Quentin, M. E., Berner, M. M. 1996, Identification of baculovirus gene that promotes *Autographa californica* nuclear polyhedrosis virus replication in a nonpermissive insect cell line. *Journal of Virology* 70: 2221-2229.

Tilakaratne, N., Hardin, S. E., Weaver, R. F. 1991, Nucleotide sequence and transcript mapping of the *Hin*dIII F region of the *Autographa californica* nuclear polyhedrosis virus. *Journal of General Virology* 72: 285-291.

Todd, J.W., Passarelli, A.L., Miller, L.K. 1996, Factors regulating baculovirus late and very late gene expression in transient-expression assays. *Journal of Virology* 70: 2307-17.

Tomalski, M. D., Miller, L. K. 1991, Insect paralysis by baculovirus-mediated expression of a mite neurotoxin gene. *Nature* 352: 82-85.

Tomalski, M. D., Miller, L. K. 1992, Expression of a paralytic neurotoxin to improve insect baculovirus as biopesticides. *Biotechnology* 10: 545-549.

Tomalski, M. D., Hutchinson, K., Todd, J., Miller, L. K. 1993a, Identification and characterization of tox2A: A mite cDNA encoding a paralytic nenrotoxin related to TxP-I. *Toxicon* 31: 319-326.

Tomalski, M., King, T. P., Miller, L. K. 1993b, Expression of hornet genes encoding venom allergen antigen 5 ih insects. *Archives of Insect Biochemistry Physiology* 22:303-313.

Treacy, M.F., All, J. N., Ghidiu G. M. 1997, Effect of ecdysteriod UDP-glucosyltransfease gene deletion on efficacy of a baculovirus against *Heliothis virescens* and *Trichoplusia ni* (Lepidoptera: Noctuidae). *Journal of Economical Entomology* 90: 1207-11214.

Van Beek, der C. P. 1980, On the origin of the polyhedral protein of the nuclear polyhedrosis virus of *Autographa californica*. Mededelingen Landbouwhogeschool Wageningen 80,5.

Van Beek, N. A. M. and Hughes, P. R., 1998, The response time of insect larvae infected with recombinant baculovirus. *Journal of Invertebrate Pathology* 72: 338-347.

Van Meer, M. M. M., Bonning, B. C., Ward, V. K., Vlak, J. M., Hammock, B. D. 2000. Recombinant, catalytically inactive juvenile hormone esterase enhances efficacy of baculovirus insecticides. *Biological Control* 19: 191-199.

Van Oers, M. M., Vlak, J. M. 1997, The baculovirus 10-kDa protein. *Journal of Invertebrate Pathology* 70: 1-17.

Van Oers, M. M., Hu, Z., Arif, B. M., van Strien E. A., van lent J. W. M., Vlak, J. M. 1998, The single – nucleocapsid nucleopolyhedrovirus of *Buzura suppressaria* encodes a P10 protein. *Journal of General Virology* 79: 1553-1562.

Van Regenmortel M. H. V., Fauquet, C.M., Bishop, D. H. L., Carstens, E. B., Estes, M. K., Lemon S. M., Maniloff, J., Mayo, M. A., McGeoch, D. J., Pringle, C. R., Wickner, R. B. 2000, Virus Taxonomy, classification and nomenclature of viruses. Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press. San Diego/San Francisco /New York/ Boston/London/Sydney/Tokyo.

Van Strien, E. A., Zuidema, D., Goldbach, R. W., Vlak, J. M. 1992, Nucleotide sequence and transcriptional analysis of the polyhedrin gene of *Spodoptera exigua* nuclear polyhedrosis virus. *Journal of General Virology* 73:2813-21.

Van Strien, E.A. 1997, Characterisation of the *Spodoptera exigua* baculovirus genome: structural and functional analysis of a 20 kb fragment. PhD thesis. Wageningen Agricultural University, Wageningen, The Netherlands.

Van Strien, E. A., IJkel, W .F. J., Gerrits, H., Vlak, J. M., Zuidema, D. 2000, Characteristics of the transactivator gene *ie-1* of *Spodoptera exigua* multiple-nucleocapsid nucleopolyhedrovirus. *Archives of Virology* 145: 2115-2133

Vlak, J. M., Smith, G. E. 1982, Orientation of the genome of *Autographa californica* nuclear polyhedrosis virus: a proposal. *Journal of Virology* 41: 1118-1121.

Vlak, J. M., Rohrmann, 1985, the nature of polyhedrin. In "*Viral Insecticides for Biological control*"(K. Maramorosh and K.E. Sherman, Eds.), Academic Press, Orlando, pp489-544.

Vlak, J. M. 1993a, Genetic engineering of baculoviruses. In "*Opportunities for Molecular Biology in Crop production*"(D. J. Beadle *et al.*, ed.), pp11-22.

Vlak, J. M. 1993b, Genetic engineering of baculovirus for insect control. In "*Molecular Approaches to Pure and Applied Entomology*" (J. G. Oakeshott and M. J Whitten, eds). Springer Series in Experimental entomology. pp 90-127.

Vlak, J. M., Schouten, A., Usmany, M., Belsham, G. J., Klinge-Roode, E. C., Maule, A. J., van Lent, J. W. M., Zuidema D. 1990, Expression of caulifower mosic virus gene I using a baculovirus vector based upon the *p10* gene using a novel selection method. *Virology.* 179: 312-320.

Wang, H., Chen, X., Wang, H., Arif, B. M., Vlak, J. M., Hu, Z. 2001, Nucleotide sequence and transcriptional analysis of a putative basic DNA-binding protein of *Helicoverpa armigera* nucleopolyhedrovirus. *Virus Genes* 22:113-120.

Wang, S. M. 1990, Cotton-wheat intercropping. Chinese Agricultural Press, Beijing, p251.

Wang, S. M. 1996, Cotton's prospect in China. *Cotton International* 42-43.

Wang, J. W., Qi, Y. P., Huang, Y. X., Li, S. D. 1995, Nucleotide sequence of a 1446 base pair *Sal*I fragment and structure of a novel early gene of *Leucania separata* nuclear polyhedrosis virus. *Archives of Virology* 140: 2 283-2291.

Watkins, M. J., Kiyatkin, N., Hughes, D., Beadle, D., King, L. 1997, Study on the biological properties of a novel recombinant baculovirus. In "*Microbial Insecticides: Novelty or Necessity?*" *BCPC Symposium. Proceedings* No. 68, pp 279-284.

Williams, C. F., Payne, C. C. 1984, The susceptibility of *Heliothis armigera* larvae to three polyhedrosis viruses. *Annals of Applied Biology* 104: 405-412.

Wilson, L. E., Wilkinson, N., Marlow, S. A., Possee, R. D., King, L. A. 1997, Identification of recombinant baculoviruses using green fluorescent protein as a selectable marker. *BioTechniques* 22: 674-681.

Wishart, M. J., Denu, J. M., Williams, J. A., Dixon, J. E. 1995, A single mutation converts a novel phosphotyrosine binding domain into a dual-specificity phosphatase. *Journal of Biological Chemistry* 270: 26782-26785.

WIV-CAS: Wuhan Institute of Virology, Chinese Academy of Sciences, 1979. Comparison of different *Heliothis armigera* nuclear-polyhedrosis virus isolates. *Acta Virologica Sinica* 1: 38-43.

Wolfe, K. H., Sharp P. M. 1993, Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. *Journal of Molecular Evolution* 37: 441-456.

Wu, Y., Carstens, E. H. 1998, A baculovirus single-stranded DNA binding protein, LEF-3, mediates the nuclear localization of the putative helicase P143. *Virology* 247: 32-40.

Xia J. Y. 1997a, Biological control of cotton aphid (*Aphis gossypii* Glover) in cotton (inter) cropping systems in China; a simulation study. PhD thesis, Wageningen University. pp1-12.

Xia, J. Y. 1997b, Chinese bollworm outbreak: causes and strategies. *Cotton International* 95-96.

Xie T., Wang, X., Nan, P., Liu, J., Zhang, G., Wang, L. 1980, Study of *Heliothis armigera* nuclear polyhedrosis virus: the acquisition of the virus envelope. *Acta Virologica Sinica* 2: 91-101.

Yearian, W. C., Luttrell, R. G., Stacey, A. L., and Young S. Y. 1980, Efficacy of *Bacillus thuringiensis* and baculovirus heliothis- chlordimeform spray mixtures against *Heliothis* spp. on cotton. *Journal of the Georgia Entomological Society* 15: 260-271.

Zanotto, P. M., Kessing B. D., Maruniak, J. E. 1993, Phylogenetic interrelationshipe among baculoviruses: Evolutionary rates and host associations. *Journal of Invertebrate Pathology* 62: 147-162.

Zanotto, P. M. de A., Sampaio, M. J. A., Johnson, D. W., Rocha, T. L., Maruniak, J. E., 1992, The *Anticarsia gemmatalis* nuclear polyhedrosis virus polyhedrin gene region: sequence analysis, gene product and structural comparisons. *Journal of General Virology* 73: 1049-1056.

Zemskov, E. A., Kang, W., Maeda, S. 2000, Evidence for nuclear binding ability and nucleosome association of *Bombyx mori* nucleopolyhedrovirus BRO proteins. *Journal of Virology* 74: 6784-6789.

Zhang, G. 1989, Commercial viral insecticide - *Heliothis armigera* viral insecticide in China. *The IPM practitioner* 11: 13.

Zhang, G. 1994, Research, development and application of *Heliothis armigera* viral pesticide in China. *Resources and Environment in the Yangtze Valley* 3: 1-6.

Zhang, G., Zhang, Y., Ge, L., Shan, Z. 1981, The production and application of the nuclear polyhedrosis virus of *Heliothis armigera* (Hübner) in biological control. *Acta Phytophylactica Sinica* 8: 235-240.

Zhang G., Zhang Y., Gong, H., Zhou, C. 1983a, Biological properties of the nuclear polyhedrosis virus of the cotton bollworm, *Heliothis armigera. Acta Virologica Sinica* 3: 63-74.

Zhang, G., Zhang, Y., Ge, L., Dan, Z., 1983b, Rearing of *Heliothis armigera* and propagation of its nuclear-polyhedrosis virus. *Acta Virologica Sinica* 3: 113-120.

Zhang G., Wang, X., Zhang, S., Gong, H., Deng, H., and Yuan L. 1983c, Morphogenesis of the nuclear-polyhedrosis virus of the cotton bollworm, *Heliothis armigera. Acta Virologica Sinica* 3: 75-85.

Zhu, H., Cong, J. P.; Mamtora, G., Gingeras, T., Shenk, Y. 1998. Cellular gene expresion altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. *Proceeding of the National Academy of Sciences of USA* 95: 14470-14475.

SUMMARY

The single nucleocapsid nucleopolyhedrovirus (SNPV) of the bollworm *Helicoverpa armigera* has been extensively used to control this insect around the world, especially in China. However, in order to compete with chemical insecticides - mainly for speed of action -novel approaches are sought to improve the efficacy of HaSNPV either by selection of superior natural variants or by genetic engineering. Prior to the development of improved HaSNPV by genetic engineering, understanding of the structure and expression strategy of the HaSNPV genome is required. This thesis describes studies aimed at the unraveling of the genetic properties of the HaSNPV genome. Furthermore, this research can provide molecular information on the taxonomic status of baculovirus morphotypes, *i.e.* single nucleocapsid NPVs (SNPV) versus multiple nucleocapsid NPVs (MNPV), and ultimately on the phylogenetic relationship among baculoviruses in general.

The polyhedrin gene, a highly conserved gene among baculoviruses and encoding the major structural protein of viral polyhedra, was localized on the HaSNPV genome and then characterized (Chapter 2). This indicated that the HaSNPV polyhedrin has a high degree of sequence similarity to that of *H. zea* SNPV. From this preliminary analysis is appeared that SNPVs are not a separate group from the MNPVs. The position of the HaSNPV polyhedrin gene was chosen as the zero point of the circular physical map of the viral genome (Chapters 5 and 6). The polyhedrin promoter, with a typical baculovirus late transcription initiation motif, was used to drive the expression of a green fluorescent protein (GFP) and a toxin in recombinant HaSNPV (Chapter 8).

In the larval stages the enzyme ecdysteroid UDP-glucosyltransferase (EGT) catalyzes the conjugation of ecdysteroid with sugars and is involved in the prevention of molting and pupation. Baculoviruses generally encode such an enzyme, resulting in the prevention of molting of infected larvae and enhanced polyhedra production. The HaSNPV *egt* gene was located on the *Hin*d-D fragment and characterized (Chapter 3). Phylogenetic analysis of this gene confirmed that HaSNPV belongs to the Group II NPVs. To further analyze the relationship between HaSNPV and other baculoviruses, a late expression factor 2' gene (*lef*-2) was identified and characterized (Chapter 4). This gene is essential for viral DNA replication and most likely functions as a DNA primase processivity factor. This is the first *lef*-2 gene characterized in any SNPV to date. Even though *lef*-2, an essential gene, and *egt*, an auxiliary gene, most likely have been under different pressure in their evolutionary past, the phylogenetic tree of baculovirus LEF-2 appeared to be comparable in form to that of EGT. The positive correlation of the genomic location of the *lef-2* genes

relative to polyhedrin/granulin genes and the clade structure of the gene trees (*lef-2*, *egt*) suggest that genome organization and gene phylogeny represent independent parameters to study the evolutionary history of baculoviruses.

In order to study the genome organization and phylogenetic status of HaSNPV, a plasmid library of its 130.1 kb-long DNA genome was made and a detailed physical map of the viral DNA was constructed (Chapter 5). From about 45 kb of dispersed sequence data generated from the plasmid library, fifty-three putative open reading frames (ORFs) with homology to ORFs of other baculoviruses were identified and their locations on the genome of HaSNPV were determined. The basic gene content of HaSNPV appeared to be quite similar to that of AcMNPV, BmNPV, and OpMNPV (group I NPVs). However, the arrangement of the ORFs along the HaSNPV genome differed significantly from that of the Group I NPVs, which all have a highly collinear genome, or that of the granulovirus XcGV. In contrast, the genomes of HaSNPV and SeMNPV (Group II NPVs) are highly collinear, both in gene content and organization. This close relatedness between an MNPV and an SNPV is supported by the phylogeny of selected genes (Chapters 2 and 3) of these two viruses and suggests that the NPV morphotype (S or M) has only a taxonomic but not a phylogenetic meaning. Homologous regions (*hrs*), a common feature of baculovirus MNPV genomes, were also located dispersed on the HaSNPV genome suggesting that their presence in common in all NPVs.

So far, only MNPV and GV genomes have been sequenced to completion, but no SNPV genome to date. Therefore the entire HaSNPV genome sequence was determined (Chapter 7). The circular, double-stranded DNA genome contains 131,403 bp and has a G+C content of 39.1 %, the lowest value among baculoviruses to date. Of 135 potential ORFs predicted from the sequence, 115 have a homologue in other baculoviruses; twenty are unique to HaSNPV and are subject to further investigation. Upon comparison with the available genomic sequences, sixty-five ORFs were found present in all baculoviruses, and hence they are considered as 'core' baculovirus genes. The HaSNPV genome lacks a homologue of the major budded virus (BV) glycoprotein gene *gp64* of group I NPVs. Instead, a functional homologue (Ha133) of *gp64* was identified after comparison with SeMNPV. The mean overall amino acid identity of the HaSNPV ORFs was the highest with SeMNPV and LdMNPV homologues. This is in accordance with their common genome organization and confirmed, that HaSNPV together with SeMNPV and LdMNPV cluster into Group II NPVs, while AcMNPV, BmNPV and OpMNPV belong to the Group I NPVs. In this analysis GV behaved like a separate group. The clade structure based on selected genes (*lef-2* and *egt*) is further strongly support by genome trees based on all conserved ORFs together and based on gene content as well as gene order on the genomes compared.

HaSNPV and HzSNPV share many common biological features such as the same

heliothine host range (Chapter 1). Sequence analysis of the complete HzSNPV genome revealed that HaSNPV and HzSNPV have a high degree of ORF identity, which is in line with the view that they are two different isolates of the same virus species (Chapters 6 and 7). The HzSNPV genome potentially encodes 139 potential ORFs of which 135 have homologous in HaSNPV. Four ORFs are unique to HzSNPV. However, these unique ORFs are small, are always found adjacent to *hr* regions and their functionality remains to be determined. Alignment of the genome sequences indicated that overall ORFs of HzSNPV have a high degree of identity with the homologues of HaSNPV genome on nucleotide (99%) and amino acid (98%) level. The 65 baculovirus core genes among these two viruses have the lowest nucleotide substitution rate, but the *hr*s showed the highest variation. Two 'baculovirus repeat orfs' (*bro*) genes in these two viruses have the highest sequence divergence and might have a different evolutionary history.

Deletion of *egt* from the baculoviral genome has been shown to increase the speed of kill of the virus and hence to reduce the crop damage by infected insects. This approach, along with the insertion of a scorpion neurotoxin gene, was used to generate recombinant HaSNPV with potentially improved insecticidal activity. The *egt* gene was deleted from the genome and replaced by the GFP and / or by an insect-specific toxin gene, AaIT (Chapter 8). Bioassay data indicated a significant reduction in the time (LT50) required for each of the HaSNPV recombinants to kill second instar *H. armigera* larvae. The LT_{50} of the *egt* deletion recombinants was about 27% shorter than that of wild type HaSNPV. The largest reduction in LT_{50} (32%) was observed when the *egt* gene was replaced by the scorpion neurotoxin AaIT gene.

The genetic and genomic analysis presented in this thesis shows that HaSNPV and HzSNPV are different variants of the same virus species. Alignment of the known baculovirus genome sequences did not clearly show the molecular basis for the baculovirus S and M NPV morphotype. Phylogenic analysis of genes and of genome organization, such as gene content and gene order, confirmed that baculoviruses can be separated into Group I and II NPV and into a GV group. Based on the investigation of the HaSNPV genome, HaSNPV recombinants with enhanced insecticidal properties were Successfully constructed providing alternative agents to bollworm control in China and elsewhere in the world.

SAMENVATTING

Het kernpolyedervirus (HaSNPV) van de katoenrups *Helicoverpa armigera* wordt wereldwijd gebruikt om dit plaaginsect te bestrijden, op grote schaal in het bijzonder in China. Om echter te kunnen concurreren met chemische insecticiden, vooral qua werkingssneldheid, worden nieuwe wegen gezocht om de effectiviteit van HaSNPV te verbeteren, hetzij via selectie van betere natuurlijke varianten, hetzij via genetische modificatie. Alvorens een verbeterd HaSNPV te kunnen ontwikkelen via genetische modificatie, is inzicht in de structuur en expressie van het HaSNPV-genoom noodzakelijk. Dit proefschrift beschrijft onderzoek, dat erop gericht is de genetische informatie van het HaSNPV-genoom op te helderen en te analyseren. Voorts levert dit onderzoek moleculaire informatie op over de taxonomische positie van de baculovirus morfotypen, d.w.z. SNPV (enkelvoudig) versus MNPV (meervoudig ingesloten nucleocapsiden per virion), en uiteindelijk over de fylogenetische verwantschap tussen baculovirussen in het algemeen.

Het polyhedrine-gen van HaSNPV, waarvan bij andere baculovirussen sterk geconserveerde homologe genen voorkomen en dat codeert voor het matrix-eiwit van baculoviruspolyeders, werd gelokaliseerd op het HaSNPV-genoom en vervolgens (Hoofdstuk 2). Sequentie-analyse liet zien dat het HaSNPV-polyhedrine een hoge mate van overeenkomst heeft met het polyhedrine van *H. zea* SNPV. Uit deze analyse bleek voorts dat op basis van het polyhedrine de SNPV's geen aparte fylogenetische eenheid vormen ten opzichte van MNPV's. De plaats van het polyhedrine-gen op het HaSNPV-genoom werd gekozen als nulpunt van de circulaire fysische kaart van dit genoom (Hoofdstukken 5 en 6). De polyhedrine-promoter, met een typerend initiatiemotief voor transcriptie van baculovirusgenen laat na infectie, werd gebruikt om de expressie van een groen-fluorescerend eiwit (GFP) en een toxine-gen in een recombinant HaSNPV te reguleren (Hoofdstuk 8).

In de larvale stadia van insecten katalyseert het enzym ecdysteroïd UDP-glucosyltransferase (EGT) de binding van suikers met ecdysteroïden en vormt het onderdeel van het regelmechanisme tijdens vervelling of verpopping. Baculovirussen coderen voor een EGT en zorgen zo voor een verlenging van de productie van nieuw virus, omdat de insectenlarven niet vervellen of verpoppen maar blijven groeien. Het HaSNPV *egt*-gen werd gelokaliseerd op het *Hin*d-D fragment en nader gekarakteriseerd (Hoofdstuk 3). Fylogenetische analyse van dit gen bevestigde dat HaSNPV moet worden

ingedeeld in groep II NPV's. Om de verwantschap tussen HaSNPV en andere baculovirussen verder te analyseren werd het gen voor de "late expressiefactor-2" (*lef-2*) geïdentificeerd en gekarakteriseerd (Hoofdstuk 4). Dit gen speelt een rol bij de virale DNA-replicatie en fungeert hoogst waarschijnlijk als co-factor van het DNA primase. Dit is het eerste *lef2*-gen dat is gekarakteriseerd in enig SNPV. Ofschoon *lef2* en *egt* (respectievelijk essentieel en niet-essentieel voor virusvermenigvuldiging sec), onder verschillende druk hebben gestaan gedurende hun evolutionaire ontwikkeling, is de fylogenetische verwantschap van baculovirus *lef2*'s vergelijkbaar met de verwantschap tussen *egt*'s. De positieve correlatie van de genomische locatie van de *lef2*-genen ten opzichte van polyhedrine-/granuline-genen met hun onderlinge afstammingsrelatie suggereert dat de genomische organisatie en de fylogenie van genen onafhankelijke parameters kunnen zijn om de evolutiegang van baculovirussen te bestuderen.

Om de genomische organisatie en de fylogenetische positie van HaSNPV verder te bestuderen werd een plasmidenbibliotheek aangelegd van het 130,1 kb lange DNA-genoom en werd een gedetailleerde fysische kaart van het virale DNA geconstrueerd (Hoofdstuk 5). Op basis van 45 kb aan verspreid liggende sequenties, gegenereerd via de plasmidenbibliotheek, werden 53 mogelijke open leesramen (ORF's) gevonden met homologie bij ORF's van andere baculovirussen en werd hun positie op het genoom van HaSNPV vastgelegd. Het gehalte aan genen van HaSNPV bleek vergelijkbaar te zijn met dat van AcNMPV, BmNPV en OpMNPV (groep I NPV's). De volgorde echter van de ORF's op het HaSNPV-genoom verschilde in belangrijke mate van die van de groep I NPV's, die alle een vergelijkbare genoomorganisatie hebben, en van het granulovirus XcGV. Daarentegen bleken de genomen van HaSNPV en SeMNPV (groep II NPV's) een vergelijkbaar gehalte aan genen te hebben alsmede een vergelijkbare genoomorganisatie. Deze grote verwantschap tussen HaSNPV en SeMNPV wordt verder bevestigd door de fylogenetische analyse van enkele geselecteerde genen van deze twee virussen (Hoofdstukken 2 en 3). Dit suggereert dat het NPV morfotype (S of M) slechts een taxonomische en geen fylogenetische betekenis heeft. Bij baculovirus MNPV's komen verspreid over het genoom gebieden voor met veel op elkaar gelijkende basenvolgorden (*hr*'s). Dergelijke gebieden werden ook verspreid over het HaSNPV-genoom gevonden en het lijkt er dus op dat het aanwezig zijn van dit type gebieden een algemene eigenschap van NPV's is.

Tot dusverre zijn alleen basenvolgorden van MNPV- en GV-genomen volledig bepaald, maar nog niet van een SNPV-genoom. Daarom werd de totale sequentie van het HaSNPV-genoom bepaald (Hoofdstuk 7). Het circulaire, dubbelstrengige DNA-genoom bevat 131.403 basenparen en heeft een G+C gehalte van 39,1%, de laagste waarde tot dusverre bij baculovirussen gevonden. Van de 135 potentiële ORF's, voorspeld via de sequentie-analyse, hebben er 115 een homoloog ORF bij andere baculovirussen; 20 ORF's blijken uniek te zijn voor HaSNPV en derhalve interessant voor verder onderzoek.

Vergelijking met de andere beschikbare baculovirus-genoomsequenties wees uit dat homologen van 65 ORF's aanwezig zijn in alle baculovirussen. Deze ORFs kunnen wellicht beschouwd worden als kern-baculovirusgenen. Een gen voor het belangrijkste oppervlakte-eiwit van het "budded" virus (BV) van groep I NPV's, GP64, ontbreekt in het HaSNPV-genoom. In plaats daarvan werd een functionele homoloog (Ha133) geïdentificeerd op basis van vergelijking met SeMNPV. De aminozuurhomologie van HaSNPV ORF's was het hoogst met ORF-homologen bij SeMNPV en LdMNPV. Dit is in overeenstemming met hun genoomorganisatie en bevestigt dat HaSNPV samen met SeMNPV en LdMNPV tot groep II NPV's kunnen worden gerekend, en AcMNPV, BmNPV en OpMNPV tot groep I NPV's. In deze analyse gedragen GV's zich overigens als een aparte groep. De veronderstelling dat het fylogenetische dendrogram, gebaseerd op enkele geselecteerde genen (*lef2*, *egt*), een goede reflectie is van de onderlinge verwantschap van baculovirussen, wordt voorts sterk ondersteund door dendrogrammen gebaseerd op alle geconserveerde ORF's tezamen, en door die gebaseerd op het gehalte aan genen en op de volgorde waarop deze genen op de respectievelijke baculovirusgenomen voorkomen.

HaSNPV en HzSNPV hebben sterk overeenkomende biologische eigenschappen, zoals dezelfde gastheren binnen de familie van de *Heliothis*-soorten. Sequentie-analyse van het complete HzSNPV-genoom liet zien dat HaSNPV en HzSNPV een hoog gehalte aan identieke ORF's hebben en derhalve weliswaar als twee verschillende isolaten kunnen worden opgevat, maar van dezelfde virussoort (Hoofdstukken 1, 6 en 7). Het HzSNPV-genoom codeert in principe voor 139 potentiële ORF's, waarvan 135 een homoloog ORF hebben in HaSNPV. Vier ORF's zijn uniek voor HzSNPV; echter, deze unieke ORF's zijn klein en liggen vaak dicht tegen *hr*-gebieden aan en hun functionaliteit moet nog worden aangetoond. Vergelijking van de beide genoomsequenties gaf verder aan dat over het algemeen HzSNPV ORF's een zeer hoge mate van identiciteit hebben met homologen op het HaSNPV-genoom op basis van nucleotiden- (99%) en aminozuurvolgorde (98%). De baculovirus 'core' genen van beide virussen hebben het laagste aantal nucleotidensubstituties, daarentegen vertoonden de *hr*'s echter een hoge variatie. De basenvolgorden van de twee "baculovirus repeat orf's"(*bro*)–genen in HaSNPV en HzSNPV zijn het meest gedivergeerd en deze genen hebben wellicht een verschillende evolutionaire oorsprong.

Verwijdering van het *egt*-gen uit het baculovirusgenoom heeft tot gevolg dat de werkingssnelheid van het virus kan worden verhoogd, waardoor de schade aan het gewas door insecten, die met deze recombinanten zijn geïnfecteerd, sterk gereduceerd wordt. Deze benadering, tezamen met de invoeging van een schorpioenentoxine-gen, werd gebruikt om recombinante HaSNPV's te maken met in principe een verbeterde insecticide-werking. Het *egt*-gen werd verwijderd uit het HaSNPV-genoom en vervangen door een gen dat codeert voor een groen-fluorescerend eiwit (merker) en / of voor een

insect-specifiek toxine-gen (Hoofdstuk 8). Uit biotoetsen bleek dat een opvallende reductie kon worden bereikt in de tijd (LT50) die HaSNPV-recombinanten nodig hebben om larven van *H. armigera* te doden. De LT50 van de *egt* deletiemutanten was ongeveer 27% korter dan die van wild type HaSNPV. De grootste reductie in LT50 (32%) werd gevonden wanneer het *egt*-gen werd vervangen door het gen dat codeert voor een schorpioenentoxine.

De genetische en genomische analyses van HaSNPV en HzSNPV, gepresenteerd in dit proefschrift, laten zien dat deze virussen varianten zijn van dezelfde virussoort, evenwel met een karakteristieke genoomorganisatie. Vergelijking van genoomsequenties van bekende baculovirussen met HaSNPV geeft nog geen duidelijk beeld van wat de moleculaire basis voor het S- of M-morfotype van NPVs is. Fylogenetische analyse van genen en van genoomorganisatie, zoals genengehalte en genenvolgorde, bevestigde dat baculovirussen opgedeeld kunnen worden in groep I en groep II NPV's en in een GV-groep. Gebaseerd op het onderzoek van het HaSNPV-genoom werden met succes recombinanten gereconstrueerd met verbeterde insecticide-eigenschappen. Deze recombinanten kunnen als alternatief worden toegepast voor chemische middelen ter bestrijding van de katoenrups in China en elders in de wereld.

总 结

棉铃虫单核衣壳核型多角体病毒(HaSNPV)已广泛用于控制棉铃虫。利用基因工程技术可改良病毒的杀虫性能，提高病毒杀虫剂与化学农药的竞争能力。对HaSNPV基因组结构和功能进行深入的研究是对其进行遗传改良的前提。本论文旨在研究HaSNPV分子遗传学，并在此基础上利用生物技术构建重组HaSNPV，获得杀虫性能改良的重组棉铃虫病毒杀虫剂。同时通过比较HaSNPV与其它杆状病毒基因组的结构，揭示SNPV和MNPV的分子基础和进化关系，研究病毒的分子分类及分子进化。

第二章首先介绍了HaSNPV的多角体蛋白基因。多角体蛋白是多角体的主要组成成分，是杆状病毒中高度保守的基因。由于其为病毒复制的非必须基因而且具有强大的启动子而广泛用于杆状病毒的基因工程研究。该章报道了HaSNPV多角体基因的定位克隆。序列分析表明HaSNPV多角体蛋白基因与HzSNPV多角体蛋白基因具有极高的同源性。本章为HaSNPV的基因组研究以及重组病毒的构建奠定了基础。

杆状病毒编码一种蜕皮激素葡萄糖糖基转移酶(EGT)。该酶催化蜕皮激素与葡萄糖基的结合而使蜕皮激素失活，从而阻断被感染的宿主昆虫的脱皮及化蛹，延长昆虫的成活时间，增加子代病毒的产量。第三章报道了HaSNPV的*egt*基因及其结构特点。EGT分子进化分析表明HaSNPV属于杆状病毒NPV组II。第四章分析了HaSNPV的晚期表达因子-2(*lef*-2)基因，以进一步确定HaSNPV在杆状病毒中的分类地位。这是首次在SNPV发现*lef-2*基因。*lef*-2可能通过与引物酶结合而参于病毒DNA复制。研究表明，虽然必须基因*lef*-2与非必须基因*egt*在进化史上具有不同的功能选择压力，但LEF-2 进化树与EGT进化树具有相似的结构。另外，不同杆状病毒*lef*2基因在基因组上的相对位置与基因进化树结构的共线性揭示基因组结构及基因可能作为杆状病毒分子进化分析的两个独立参数。

为了深入研究HaSNPV的基因组结构及分类地位，首先构建了HaSNPV基因组DNA质粒文库及精细的的物理图谱(第五章)。利用所构建的质粒文库，测定了基因组中约45 kb的序列，鉴定了53个杆状病毒的同源基因。结果表明虽然HaSNPV具有与AcMNPV, BmNPV及OpMNPV(NPV组I)具有相似的基因组成，但基因在基因组上的排列与杆状病毒NPV组I和GV有显著的不同。相反，HaSNPV无论是基因组成还是基因在基因组上的排列分布均与SeMNPV和LdMNPV具较高的相似性。MNPV和SNPV这种相关性，说明杆状病毒的形态特征单粒包埋(S)和多粒包埋(M)仅仅为分类学特征，并不能作为进化分析的指标。这一结论也得到所选择基因的分子进化分析的支持。

到目前为止，仅有MNPV和GV的全序列被测定，尚无SNPV基因组全序列的报导。第六章分析了HaSNPV基因组全序列。HaSNPV双链环状DNA基因组为131,403 bp，G+C含量为39.1 %，是目前已报道的杆状病毒中G＋C含量最低的。计算机分析发现了135可能的开放阅读框(ORF)，其中115 ORF在其它杆状病毒中存在同源基因，20个ORF为HaSNPV的特有基因。通过与已有的基因组序列的比较发现65 ORF在杆状病毒基因组中完全保守，可能为杆状病毒的"核心"基因。HaSNPV基因组缺乏BV的主要糖蛋白基因*gp64*，但具有*gp64*功能同源基因。与其它杆状病毒相比，HaSNPV的ORF与SeMNPV和LdMNPV的ORF的氨基酸序列普遍具有较高的同源性，这进一步证明HaSNPV，SeMNPV及LdMNPV属于NPV组II，AcMNPV，BmNPV和OpMNPV属于NPV组I，而GV处于一个独立的进化分支。杆状病毒的这种分子进化关系进一步得到所有保守基因联合进化树及基因组成和基因分布的基因组进化树的证实（第九章）。

HaSNPV和HzSNPV的ORF具有极高的同源性(第1, 6, 7章)，同时它们还具有相似的生物学特性，如相同的宿主范围等。HzSNPV基因组全序列分析表明HaSNPV和HzSNPV为同一病毒种的两个不同的变种（第七章）。HzSNPV基因组可能编码139 ORF，其中135 ORF与HaSNPV同源。4 ORF为HzSNPV特有，但这些ORF均较小并与同源重复区（*hr*）邻近。序列比较分析表明HzSNPV绝大部分ORF与HaSNPV具有极高核苷酸(99%)及氨基酸(98%)同源性，其中，在所有杆状病毒中保守的基因的核苷酸取代率最低，而*hr*区域为高可变区。两个病毒中的杆状病毒重复ORF基因(*bro*)具有最高的异质性，可能具有不同的进化历程。

研究表明删除杆状病毒*egt*可提高杆状病毒的杀虫速度减少感病昆虫的危害。缺失*egt*及表达蝎子的神经毒素两种不同方法用于构建重组HaSNPV病毒杀虫剂，以提高其杀虫速度（第八章）。以二龄棉铃虫进行生物测定结果表明所获重组HaSNPV的杀虫速度均明显缩短。*Egt* 缺失的重组病毒的LT_{50}比野生型病毒减少27%。最佳改良效果为*egt*基因缺失并表达昆虫特异性神经毒素基因（AaIT）的重组HaSNPV，其LT_{50}比野生型病毒减少32%。

本论文的遗传学及基因组分析表明HaSNPV和HzSNPV为同一病毒种的两个不同变种。基因组序列分析并未清楚揭示杆状病毒不同表型S和M的分子基础。基因及基因组的基因组成和基因的排列分布的进化分析进一步证实杆状病毒分为NPV和GV，而NPV进一步分为组I和组II。在对HaSNPV基因组研究的基础上成功的构建了杀虫性能改良的重组HaSNPV，为棉铃虫的控制提供了新的途径。

LIST OF PUBLICATIONS

Chen X., Hu Z., Jehle, J.A, Zhang, Y., Vlak, J. M. 1997. Analysis of the ecdysteroid UDP-glucosyltransferase gene of *Heliothis armigera* single-nucleocapsid baculovirus. *Virus Genes* 153: 219-225.

Chen, X., Hu, Z. H., Vlak, J. M. 1997. Nucleotide sequence analysis of the polyhedrin gene of *Heliothis armigera* single nucleocapsid nuclear polyhedrosis virus. *Virologica Sinica* 12: 346-353*.

Hu, Z. H., Arif, B. M., Sun, J. S., **Chen, X.,** Zuidema, D., Goldbach, R. W., Vlak, J. M. 1998. Genetic organization of the *Hin*dIII-I region of the single-nucleocapsid nucleopolyhedrovirus of *Buzura suppressaria*. *Virus Research* 55: 71-82.

Hu, Z. H., Arif, B. M., Jin, F., Martens, J. W., **Chen, X.** W., Sun, J. S., Zuidema, D., Goldbach, R. W., Vlak, J. M. 1998. Distinct gene arrangement in the *Buzura suppressaria* single-nucleocapsid nucleopolyhedrovirus genome. *Journal of General of Virology* 79: 2841-51.

Chen, X., IJkel, W. F., Dominy, C., de Andrade, Zanotto, P. M., Hashimoto, Y., Faktor, O., Hayakawa, T., Wang, C. H., Prekumar, A., Mathavan, S., Krell, P. J., Hu, Z. H, Vlak, J. M. 1999. Identification, sequence analysis and phylogeny of the *lef*-2 gene of *Helicoverpa armigera* single-nucleocapsid baculovirus. *Virus Research* 65: 21-32.

Luo, B., Wang, H., **Chen, X.,** Sun, X., Wang, H., Peng, H., Arif, B. M., Vlak, J. M. 1999. Genetic organization off the *Bam*HI-H region of the single nucleocapsid nucleopolyhedrovirus of *Buzura suppressaria* (BusuNPV). *Virologica Sinica* 14: 333-342*.

Chen, X., Sun, X., Hu, Z., Li, M., O'Reilly, D. R., Zuidema, D., Vlak, J. M. 2000. Genetic engineering of *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus as an improved pesticide. *Journal of Invertebrate Pathology* 76: 140-1466.

Chen, X., Li, M., Sun, X., Arif, B. M., Hu, Z., Vlak, J. M. 2000. Genomic organization of *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus. *Archives of Virology* 145: 2539-2555.

Wang, H., **Chen, X.**, Wang, H., Arif, B. M., Vlak, J. M., Hu, Z. 2001. Nucleotide sequence and transcriptional analysis of a putative basic DNA-binding protein of *Helicoverpa armigera* nucleopolyhedrovirus. *Virus Genes* 22: 113-120.

Chen, X., IJkel, W. F. J., Tarchini, R., Sun, X., Sandbrink, H., Wang, H., Peters, S., Zuidema, D., Lankhorst, R. K., Vlak, J. M., Hu, Z. 2001. The sequence of the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus genome. *Journal of General of Virology* 82: 241-257.

Herniou, E. A., Luque, T., **Chen, X.**, Vlak, J. M., Winstanley, D., Cory, J. S., O'Reilly, D. R. 2001 Baculovirus phylogenies based on gene sequences, gene order and gene content. *Journal of Virology* 75: 8117-8126.

Sun, X., **Chen, X.**, Zhang, Z., Wang, H., Bianchi, F. J. J. A., Peng, H., Vlak, J. M., Hu, Z. H. 2001. Bollworm response to release of genetically modified *Helicoverpa armigera* nucleopolyhedroviruses in cotton. *Journal of Invertebrate Pathology.* Accepted.

Chen, X., Zhang, W-J., Wong, J., Chun, G., Lu, A., McCutchen, B. F., Presnail, J. K., Herrmann, R., Dolan, M., Tingey, S., Hu, Z., Vlak, J. M. 2001. Comparative analysis of the complete genome sequences of *Helicoverpa zea* and *H. armigera* single-nucleocapsid nucleopolyhedrovirus. *Journal of General Virology.* Accepted.

* in Chinese

ACKNOWLEDGEMENTS

By joining the cooperative project between professor Just M. Vlak and professor Hu Zhihong on the bollworm baculoviruses I had the first chance to visit Wageningen in 1996. This visit led me to the track of my PhD study. Traveling between China and the Netherlands for many years enriched my professional life and allowed me to broaden the horizons of my personal life too. This venture has been one of the most important experiences in my life.

I am grateful to my promotor professor Just Vlak for his scientific guidance and our indepth discussions, which elevated the standards of my research and my life. His critical comments on all chapters have been of great value to me. I am indebted to him and his wife Ellen for their patience with me and for spending evenings and weekends to revise my papers and thesis. The hospitality and friendship of the Vlak family made my stay in the Netherlands very enjoyable. I thank professor Rob Goldbach for his interest in my research, which has been a constant source of inspiration.

My co-promotor professor Hu Zhihong's (Rose) day to day encouragement was indispensable for the successful completion of this thesis. Rose, also as a friend and colleague, I will never forget how we discussed so many things ranging from science, philosophy to personal life. Thank you very much for taking over all my responsibilities in China, when I was in the Netherlands. I also appreciate very much the friendship of Yiwu .

The work presented in this thesis would not have been completed without the help of many international colleagues. My thanks first go to Dr Johannes Jehle (Zentrum fuer Gruene Gentechnik, Neustadt a/d Weinstrasse, Germany). I still remember how you helped me to analyze my first RE gel and to construct my first phylogenetic tree. I am specifically grateful to David O'Reilly (Imperial College of Science and Medicine, London, UK) for his help in the analysis of EGT activity of the recombinant HaSNPVs and for the cooperation on baculovirus genome phylogeny. I am grateful to professor Peter Krell (University of Guelph, Guelph, Canada) and Dr Basil Arif (Great Lakes Forestry Centre, Sault Ste. Marie, Canada) for their continuous interest in my research and their valuable comments and suggestions on the thesis. I thank Dr Martijn Huynen (EMBL, Heidelberg, Germany) for the helpful discussion and the generation of the baculovirus genome tree. The cooperation with Dr Albert Lu (Dupont Agricultural Products, USA) on the HzSNPV genome sequence project was highly appreciated.

I wish to express my sincerest thanks to my friends and colleagues in the Laboratory of Virology in Wageningen. Special thanks go to Douwe Zuidema for his scientific input

and continuous support, to Monique van Oers for her helpful discussions and friendship, and to Magda, Els, and Angela for their technical help. I also wish to acknowledge and thank everyone in the baculovirus group: Mariëlle van Hulten, Wilfred IJkel, Felix Bianchi, Gorben Pijlman, Marcel Westenberg, Stephen Kaba, Xiao Jiang, Yang Kai and my former colleagues Lilian Colbers, René Broer and Jacco Heldens. My appreciation also goes to colleagues and friends outside the baculovirus group, in particular to Wout Rozeboom, Thea van Bemmel, Marcel Prins and Dick Peters. I thank all the colleagues, the baculo's, the tospo's, the whispo's and the lento's for the good times for so many years. I will miss the beer times and coffee breaks.

I am grateful to my home institute: the Wuhan Institute of Virology, Chinese Academy of Sciences, in Wuhan for the solid support to advance my scientific development and for granting me such an extended stay in the Netherlands. Special thanks are due to the directors of the institute, professor Hu Zhihong, Li Xinge, Gong Hanzhou and Yuan Zhiming and previous director professor He Tianfu, for their constant support and understanding. I owe deep gratitude to professor Zhang Xianen for his encouragement, understanding and moral support during all the years of my scientific career. Many of my colleagues at the institute provided help, in one way or another, to my research. My profound appreciation thus goes to professor Xie Tianen, Jin Feng, Zhang Yinlian, Peng Huiying, Zhang Youqi, Sun Xiulian, Wang Hanzhong, Shi Zhengli, Li Tianxian, Liang Li and Li Mei. I thank all the students in our lab in Wuhan for their hard work and good results in my physical absence.

I thank all my Chinese friends in Wageningen for their help and the enjoyable time in the Netherlands. I would like to specifically acknowledge in this regard the Huang Jiang family, the Chunming family, Fuyi, Hugui, Xu Han, Zhang Lei, Zhang Peng, the Liping family and the Xinyou family. My sincere indebtedness is extended to Mr. Tom Schiphorst and Ms Zhang Liru for their kindness over the years.

Finally, I would like to thank my dear family for a life-long support. I express my sincerest gratitude and love to my wife, Rongrong, and my son, Chencheng. Thank you very much for your moral support, frequent encouragement, and your understanding and patience during this venture. Quality time together is ahead! I am also very indebted to my parents and to my parents-in-law, for helping to take care of Chencheng.